

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 27, 2003, 03:14:59 ; Search time 2282 Seconds

(without alignments)
13620.419 Million cell updates/sec

Title: US-09-675-509-3

Perfect score: 1068

Sequence: 1 atgtccatcaccacaaagac.....gtgtctgagaccatccctc 1068

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : GenEmbl:*

1: gb_ba:*
2: gb_hcg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_hcg_hum:*
31: em_hcg_inv:*
32: em_hcg_other:*
33: em_hcg_mus:*
34: em_hcg_pln:*
35: em_hcg_rtd:*
36: em_hcg_man:*
37: em_hcg_vit:*
38: em_sy:*
39: em_higo_hum:*
40: em_higo_mus:*
41: em_higo_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	54.2	5.1	7218	6 166494	166494 Sequence 14
2	52.6	4.9	1141	6 AX083744	AX083744 Sequence
3	50	4.7	4197	3 AF163835	AF163835 Dictyostel
4	48.4	4.5	441	1 RP282600	282600 R. prowazeki
5	47.8	4.5	195829	9 AL353689	AL353689 Human DNA
6	46.6	4.4	146568	2 AC022651	AC022651 Homo sapi
7	46.6	4.4	165236	9 AC024255	AC024255 Homo sapi
8	46	4.3	1141	6 AX083744	AX083744 Sequence
9	46	4.3	174253	2 AC068590	AC068590 Homo sapi
10	46	4.3	187254	2 AC124649	AC124649 Homo sapi
11	46	4.3	195912	2 AC103465	AC103465 Rattus no
12	45.4	4.3	114741	2 AC124954	AC124954 Medicago
13	44.6	4.2	282610	1 RPXX01	AJ235270 Rickettsi
14	44.6	4.2	304050	1 AP004829	AP004829 Scaphyloc
15	44.6	4.2	307750	1 AP003136	AP003136 Scaphyloc
16	44.4	4.2	138016	9 CNS01DTP	AL132827 Human chr
17	44.4	4.2	171942	2 AC021715	AC021715 Homo sapi
18	44.4	4.2	175948	2 AC019192	AC019192 Homo sapi
19	44.2	4.1	77096	2 AC116030	AC116030 dictyoste
20	44.2	4.1	148566	2 AC097671	AC097671 Rattus no
21	44.2	4.1	181254	2 AP000825	AP000825 Homo sapi
22	44.2	4.1	181589	9 AP000868	AP000868 Homo sapi
23	44	4.1	19089	2 AC026872	AC026872 Homo sapi
24	44	4.1	174603	9 AP003096	AP003096 Homo sapi
25	44	4.1	188250	9 AP001788	AP001788 Homo sapi
26	43.8	4.1	14461	6 AX256438	AX256438 Sequence
27	43.4	4.1	146383	2 AC116367	AC116367 Oryza sat
28	43.4	4.1	168698	9 AC068138	AC068138 Homo sapi
29	43.2	4.0	29520	3 CEM06F12	Z83244 Caenorhabdi
30	43.2	4.0	97683	2 AC116548	AC116548 Dictyoste
31	43.2	4.0	104025	9 AL138679	AL138679 Human DNA
32	43.2	4.0	104485	2 AC097864	AC097864 Rattus no
33	43.2	4.0	158840	9 AC091577	AC091577 Homo sapi
34	43.2	4.0	165018	9 AC079173	AC079173 Homo sapi
35	43.2	4.0	172657	9 AC015963	AC015963 Homo sapi
36	43.2	4.0	189148	2 AC068262	AC068262 Homo sapi
37	43	4.0	110000	2 AC113004	AC113004 Mus muscu
38	43	4.0	122515	2 AC120261	AC120261 Rattus no
39	43	4.0	348650	1 AP003364	AP003364 Scaphyloc
40	42.8	4.0	5201	1 AP011104	AJ011104 Apple pro
41	42.8	4.0	73039	2 AC116957	AC116957 Dictyoste
42	42.8	4.0	110000	2 AC125102_2	Continuation (3 of
43	42.8	4.0	124191	9 AC005153	AC005153 Homo sapi
44	42.8	4.0	162925	2 AP004461	AP004461 Oryza sat
45	42.8	4.0	175905	2 AC127263	AC127263 Mus muscu

ALIGNMENTS

RESULT 1
LOCUS 166494
DEFINITION Sequence 14 from patent US 5670367.
ACCESSION 166494
VERSION 166494.1 GI:2724471
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 7218)
AUTHORS Dornier, F., Scheitlinger, F. and Falkner, F. Gunter.
TITLE Recombinant fowlpox virus
JOURNAL Patent: US 5670367-A 14 23-SEP-1997;
FEATURES Location/Qualifiers

```

source          1..7218
BASE COUNT      1944 a    1491 c    1406 g    1929 t    368 others
ORIGIN
Query Match     5.1%; Score 54.2; DB 6; Length 7218;
Best Local Similarity 14.5%; Pred. No. 0.38;
Matches 83; Conservative 220; Mismatches 268; Indels 0; Gaps 0;

Oy 142 ACCGAAATTGATGTTCACAGTGAATGCTACTCTTCAAAAGCTCCACAGATGTTTCCAACT 201
Db 1021 ACAGAATTATATCCGAGCTTGCGTAGGTGCAAGGAGCTTGGAATTTTTTTTTTTTTTT 1080
Oy 202 GATGACATTTCTCCTCATATCTGGTCTCTTGGGAGGTGCAAGATTTGATGATCA 261
Db 1081 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 1140
Oy 262 TTGGTTCTGGTGTTACTGATGATTTGCATAGTTTGTTCCTCAAGTCCCTGTCAAT 321
Db 1141 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 1200
Oy 322 GGTCGCGTTAAGTTCACCACAATCTGTGCTCAAACCTTTTATGTCTCACCAAT 381
Db 1201 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 1260
Oy 382 GGTACTCAACAAGCATCTCCCTTTAGAAATGGCTCAAAAGGTGGTATGAACAAT 441
Db 1261 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 1320
Oy 442 GTTTATCCAGATGTCCTCTTCTAGTCTTTTCACAGTTTTCGATTTGATCAACAAT 501
Db 1321 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 1380
Oy 502 CTCGATCATCATCATCAGCTGAGTGAATATCAAGGCTCTGATCTTCCACAATCTG 561
Db 1381 TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 1440
Oy 562 GACCAAGTCAACAAGATATCACTCAAAAATAATAGAACATTTGGATTCAACAGTGT 621
Db 1441 AATTCTTATCTCTTAACTACTTGCATAGATAGTATATACAGTGAAGTCCCTACATGCC 1500
Oy 622 GCCTCTCAAGAGATATATTAATCTGTGAAGCAAGTAAACCAATTTCAACTATAT 681
Db 1501 GTTTTGTGAAACGAAATAGATGCTGTGAAGGAGTACGTACGATCAACAATCACT 1560
Oy 682 CTCGATATAGTGAAGTANGTGAAATTA 712
Db 1561 TTCATATTTAGAAATATATGATGTAATAAATA 1591

RESULT 2
AX083744 1141 bp DNA linear PAT 28-FEB-2001
LOCUS AX083744
DEFINITION Sequence 22 from Patent WO011061.
VERSION AX083744 GI:13185472
KEYWORDS
SOURCE
ORGANISM synthetic construct.
REFERENCE artificial sequences.
AUTHORS Kunst, L. and Clemens, S.
TITLE Regulation of embryonic transcription in plants
JOURNAL Patent: WO 011061-A 22 15-FEB-2001;
UNIVERSITY OF BRITISH COLUMBIA (CA)
FEATURES
source location/Qualifiers
1..1141
/promoterm="synthetic construct"
/promoter "note="sequence of A.T., L.A., and B.N. PAEI
promoters"
1..1141
/promoter "note="sequence of A.T., L.A., and B.N. PAEI
promoters"
1..1141
/promoter "note="sequence of A.T., L.A., and B.N. PAEI
promoters"

```

```

ORIGIN
Query Match          4.9%; Score 52.6; DB 6; Length 1141;
Best Local Similarity 11.0%; Pred. No. 1.1;
Matches 53; Conservative 196; Mismatches 234; Indels 0; Gaps 0;

Oy 43 TATCTTCCTTCTGGAAATGAAAAATGGCAACAACATTTAAATGATCAATTTGATCAAGGAT 102
    ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 182 HMMNNNGEITWTVMVRYKTDSDMSBKNNYGMBMXKMSYIVTYTMMWDMCKRKYRW 241

Oy 103 GTTTTGGCAACTCAGGTTTCCGATATATATATATGCAATATACCGAATTTGATTGTAAGT 162
    ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 242 VRTGRMENVAVWABTAHRRRYNNGWTBAMAYRRMTMNNNNNNNNKAMCKRAKYGMWRAB 301

Oy 163 GATGCTACTCTTCAAAAGCTTCCAGATGTTTCTCAACTGATACATTTCTCCATAT 222
    ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 302 VNSTCTTWMKSTTKYRISCMANNCRAGDANKDHKMKWSALMGVMMNNNNNNNTYKAR 361

Oy 223 CTGTGTTCTTTGGGTTGGTGTCAGAAGTTTGATGAATCATTTGTCGTGCTGTTACTGTT 282
    ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 362 HBARDDWVHSAWKMKHANAHAHYRKKMTBYRKRTVMNNNNNGTTWMKXMAWYMKMDW 421

Oy 283 GATTGCATAGCTTTGTTGTTTCTCAAGTCCCTGTCAATGTTCCGTTATGTTTCCCA 342
    ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 422 BGTVMNNNNNGRTYGYGMTKNKKMTYYKMKANNCKWABMDHKTCHNNTTWMKKKTYWNN 481

Oy 343 CAATCTCTGTCTCAAACTTTTATATGTCCTCACCAATAGTACTCAACAAGCATCTTCC 402
    |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 482 CWMKSMGTGKSHRBAALAYTYWMMWRRYAANNNNNDYMKACTGKYBVCSKMMNNYA 541

Oy 403 CTTTAGAAATTTGGCTCAAAAGTTGGTATGAACAATTTGTTATCCAGATGTGCTCT 462
    ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 542 AMYTSSNNYTSRYYRMTKNNNSWRMSDTRSBMGANAYARABHGYKMTWMBMSHTWB 601

Oy 463 TCTAGTTCTTTCAAGTTTTCGATTGTATGAACAATTAATCTCCAAATCATCATCAAGT 522
    ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 602 HBRAGAAYHMBMYBAKCHCMKAWYKAKKYAGAGSNNNNNNNNNNNNNNNNNNATCAADD 661

Oy 523 GCA 525
    ::|
Db 662 YYA 664

RESULT 3
LOCUS AF163835 4197 bp mRNA linear INV 01-MAR-2000
DEFINITION Dictyostelium discoideum cellulose synthase (dcsa) mRNA, complete cds.
ACCESSION AF163835
VERSION AF163835.1 GI:6002920
KEYWORDS Dictyostelium discoideum.
SOURCE Dictyostelium discoideum.
ORGANISM Dictyostelium discoideum.
REFERENCE 1 (bases 1 to 4197)
AUTHORS Blanton,R.L., Fuller,D., Iranfar,N., Grimson,M.J. and Loomis,W.F.
TITLE The cellulose synthase gene of Dictyostelium
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (5), 2391-2396 (2000)
MEDLINE 20160962
PUBMED 10681463
REFERENCE 2 (bases 1 to 4197)
AUTHORS Blanton,R.L., Fuller,D., Iranfar,N., Grimson,M.J. and Loomis,W.F.
TITLE Direct Submission
JOURNAL Submitted (29-JUN-1999) Center for Molecular Genetics, UCSD, 9500 Gilman Drive, La Jolla, CA 92093, USA
FEATURES
source
1. 4197
/organism="Dictyostelium discoideum"
/strain="AX4"
/db_xref="taxon:44689"
1. 4197
/gene="dcsA"
806..3985

```


COMMENT

CB10 15A, UK. E-mail enquiries: humquery@sanger.ac.uk
 requests: clonerequest@sanger.ac.uk
 On Oct 21, 2001 this sequence version replaced gi:15983876.
 During sequence assembly data is compared from overlapping clones.
 Where differences are found these are annotated as variations
 together with a note of the overlapping clone name. Note that the
 variation annotation may not be found in the sequence submission
 corresponding to the overlapping clone, as we submit sequences with
 only a small overlap as described above.
 This sequence was finished as follows unless otherwise noted: all
 regions were either double-stranded or sequenced with an alternate
 chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such
 as compressions and repeats; all regions were covered by at least
 one plasmid subclone or more than one M13 subclone; and the
 assembly was confirmed by restriction digest. The following
 abbreviations are used to associate primary accession numbers given
 in the feature table with their source databases: Em: EMBL, Sw:
 SWISSPROT; Tr: TrEMBL; Wp: WormPep; information on the WormPep
 database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence
 was generated from part of bacterial clone contigs of human
 chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping
 Group. Further information can be found at
<http://www.sanger.ac.uk/HGP/Chr1>
 RP5-1087E8 is from the library RPCI-5 constructed by the group of
 Plexer de Jong. For further details see
<http://www.chori.org/bacpac/home.htm>
 VECTOR: pCYPAC2
 This sequence is the entire insert of clone RP5-1087E8 The true
 left end of clone RP1-1B20 is at 155425 in this sequence. The true
 right end of clone RP11-375H24 is at 75234 in this sequence.
 Location/Qualifiers

FEATURES

source

1. 195829
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="1"
 /map="q42.11-42.2"
 /clone="RP5-1087E8"
 /clone_1lb="RPCI-5"
 13995..14136
 /note="Sequence from overlapping clone RP11-375H24
 (AL391628). Assembly confirmed by restriction digest."
 17759
 /note="Tandem repeat. Forced join. Gap size estimated to
 be approximately 250bp by restriction digest data."
 22581
 /note="Tandem repeat. Forced join. Gap size estimated to
 be approximately 600bp by restriction digest data."
 38361
 complement (38872..39056)
 /note="Sequence from overlapping clone RP11-375H24
 (AL391628). Assembly confirmed by restriction digest."
 39840..39867
 /note="Single clone region. Assembly confirmed by
 restriction digest data."
 195130..195532
 /note="Sequence from overlapping clone RP11-1B20
 (AL451047). Assembly confirmed by restriction digest."
 BASE COUNT 57449 a 40631 c 39653 g 58096 t
 ORIGIN

Query Match 4.5%; Score 47.8; DB 9; Length 195829;
 Best Local Similarity 53.5%; Pred. No. 3.9;
 Matches 100; Conservative 0; Mismatches 87; Indels 0; Gaps 0;
 QY 551 CACAATCTGTGACCAAGTCAACAGATATCACTCAAAATATAGAACCATTTGGATT 610
 Db 135701 CACCATATGTTGTATTATTATACAAATATATATTCCTAAACATGAGAGATTTTAGGTT 135760
 QY 611 CAACAGTTTGCTCTCAAGAGATATATTAACTGTGAAGCAAGTAACCAATTT 670
 Db 135761 TAAATAAATTTGAGAGTCACATAATTTTATGAGAGAAATTAATTAATGAAATATTATT 135820

QY 671 CAACACTATATGTCGATATATGGAAGTATGTTGTAATTAAGATATCATGAGATC 730
 Db 135821 AATACATATATATCTACTTATATGACACAGCAGAGAAAACAGGCTATATTAACCTC 135880
 QY 731 AACCAAT 737
 Db 135881 ATCACTA 135887

RESULT 6

AC022651

LOCUS

DEFINITION

AC022651

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

AC022651 146566 bp DNA linear HTG 01-MAR-2000
 Homo sapiens clone RP11-28H7, WORKING DRAFT SEQUENCE, 12 unordered
 pieces.
 AC022651 GI:7139694
 HTG; HTGS_PHASE1; HTGS_DRAFT.
 Homo sapiens.
 Homo sapiens
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 146568)
 Birren,B., Linton,L., Nusbaum,C. and Lander,E.
 Homo sapiens, clone RP11-28H7
 Unpublished
 2 (bases 1 to 146568)
 Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
 Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
 Boguslavsky,L., Boukhalter,B., Brown,A., Burkett,G., Castle,A.,
 Choepel,Y., Colangelo,M., Collins,S., Collamore,A., Cooke,P.,
 Dearlellano,K., Dewar,K., Domino,M., Doyle,M., Fenster,J.,
 Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
 Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
 Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
 Landers,T., Lebecky,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
 MacDonald,P., Margis,N., McEwan,P., McGarr,A., McKernan,K.,
 McPherson,R., Meldrum,J., Menais,L., Morrow,J., Naylor,J.,
 Norman,C.H., O'Connor,T., O'Donnell,P., Oliver,T.M., Peterson,K.,
 Pierre,N., Plesni,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
 Roy,A., Santos,R., Severy,P., Spencer,B., Strange-Thomann,N.,
 Stojanovic,N., Sudramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
 Turrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
 Zimmer,A. and Zody,M.
 Direct Submission
 Submitted (06-FEB-2000) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Mar 1, 2000 this sequence version replaced gi:6910657.
 All repeats were identified using RepeatMasker:
<http://ftp.genome.washington.edu/RW/RepeatMasker.html>
 ----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: <http://www-seg.wi.mit.edu>
 Contact: sequence submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: L4767
 Center clone name: 28_H.7
 ----- Summary Statistics
 Sequencing vector: M13; M77815, 100% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.960731
 Consensus quality: 140377 bases at least Q40
 Consensus quality: 143582 bases at least Q30
 Consensus quality: 144644 bases at least Q20
 Insert size: 145000; agarose-fp
 Insert size: 145468; sum-of-contrigs
 Quality coverage: 4.8 in Q20 bases; agarose-fp
 Quality coverage: 4.8 in Q20 bases; sum-of-contrigs

 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 12 contigs. The true order of the pieces

* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved. 2283: contig of 2283 bp in length
1
2284 2383: gap of 100 bp
2384 9073: contig of 6690 bp in length
9074 9173: gap of 100 bp
9174 16904: contig of 7731 bp in length
16905 17004: gap of 100 bp
17005 25601: contig of 8597 bp in length
25602 25701: gap of 100 bp
25702 34479: contig of 8778 bp in length
34480 34579: gap of 100 bp
34580 41772: contig of 7193 bp in length
41773 41872: gap of 100 bp
41873 55432: contig of 13560 bp in length
55433 55532: gap of 100 bp
55533 68980: contig of 13448 bp in length
68981 69080: gap of 100 bp
69081 84607: contig of 15527 bp in length
84608 84707: gap of 100 bp
84708 100925: contig of 16218 bp in length
100926 101025: gap of 100 bp
101026 115546: contig of 14521 bp in length
115547 115646: gap of 100 bp
115647 146568: contig of 30922 bp in length.
Location/Qualifiers
1. 146568
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="RP11-28H7"
/clone_1ib="RP11-11 Human Male BAC"

FEATURES

source

misc_feature
1..2283
/note="assembly_fragment"
misc_feature
2384..9073
/note="assembly_fragment"
misc_feature
9174..16904
/note="assembly_fragment"
misc_feature
17005..25601
/note="assembly_fragment"
misc_feature
25702..34479
/note="assembly_fragment"
misc_feature
34580..41772
/note="assembly_fragment"
misc_feature
41873..55432
/note="assembly_fragment"
misc_feature
55533..68980
/note="assembly_fragment"
misc_feature
69081..84607
/note="assembly_fragment"
misc_feature
84708..100925
/note="assembly_fragment"
misc_feature
101026..115546
/note="assembly_fragment"
misc_feature
115647..146568
/note="assembly_fragment"
clone_end:SP6
vector_end:left
BASE COUNT 44763 a 27525 c 28254 g 44924 t 1102 others
ORIGIN

Query Match 4.4%; Score 46.6; DB 2; Length 146568;
Best Local Similarity 48.0%; Pred. No. 7.1;
Matches 133; Conservative 0; Mismatches 144; Indels 0; Gaps 0;

QY 584 CTCAAAATATAGAACCATTTGGATTCAACAGTTGGCTCTCAAGAATATATTTA 643
Db 118836 CACGAAATGCACATCAAACTCAATGAGATATCATCTCAACCCCACTTAAATGCTTA 118895

QY 644 ACTCTGTAAAGCAAGGTAAACCAATTTCAACTACTATGTCTGATATAGTAAAGTATGT 703
Db 118896 TATCCAAAGACAGCGCAATTAACAATGCTGGTGAATGTGGAAAGGAAACCTCGT 118955
QY 704 GTAAATTAAGGATTCATCATGAGATCAACATATCAATGTTCACATCTTGTGACTCTG 763
Db 118956 ACATATTAGTGAGAAATGTAAATTAGCACAACTACATGAGATATCTTGTGACAGTTCTT 119015
QY 764 ATAGCATATCGTTTATACGATGTTTGGCTTGAATTCATTTGTTGTGATGAAGAAAGC 823
Db 119016 CAAAACCTTAAATACACTGTGATATGATGCGAGAAATCCACTGCTGGATATATACC 119075
QY 824 AAAAGTTGCTGTGAAGTTATCAAGATTTATTCAG 860
Db 119076 AAAAGAAAGCAATTCAGTGTATCAAGATATCTMC 119112

RESULT 7

AC024255/c

LOCUS

DEFINITION

165236 bp DNA linear PRI 30-NOV-2000
Homo sapiens 12 BAC RP11-31A23 (Roswell Park Cancer Institute Human

BAC library) complete sequence.

AC024255

AC024255 GI:11465004

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-ouman,F.R., Allen,C.,
Alshrocks,S.L., Amaralunge,H.C., Are,J.R., Banke,T., Barberia,J.,
Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Buck,J.,
Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Bunay,C.,
Burich,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,R.,
Chen,Z., Chiu,D., Chowdhury,I., Christopoulos,C., Cleveland,C.D.,
Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
Dem,A.L., Ding,Y., Din,H.H., Douthwaite,K.J., Draper,H.,
Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
Elhaj,C., Emerling,S., Escotto,M., Falls,T., Ferraguto,D.,
Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J.,
Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W.,
Guaracane,P., Hale,S., Hamilton,K., Han,J., Harris,C., Harris,K.,
Hart,M., Havlak,P., Hawes,A., Hernandez,J., Hernandez,O.,
Hodgson,A., Hognes,M., Holloway,C., Hollins,B., Homsi,F.,
Howard,S., Huber,J., Huylk,S., Hume,J., Ioshikhes,I., Jackson,L.E.,
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
Karlsone,E., Kelly,S., Khan,U., King,L., Kovach,J., Kovar,C.,
Kratovic,J., Kureishi,A., Landry,N., Leal,B., Lee,E., Lewis,L.C.,
Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
LouiSeged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
Ma,J., Maheshwari,M., Mapua,P., Marondel,I., Martin,R.,
Martindale,A., Martinez,E., Massey,E., Mawhney,E., McLeod,M.P.,
Meador,M., Mei,G., Merscher,S., Metzger,M., Miller,A., Miner,G.,
Miner,Z., Mitchell,T., Mohabbat,K., Montgomerly,K.T., Morgan,N.,
Morris,S., Moser,M., Neal,D., Nelson,D., Newton,J., Newton,M.,
Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokwenwo,S.,
Ogun,M., Okunolu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
Peery,J., Perez,L., Peters,L., Pickens,R., Prims,E., Pu,L.L.,
Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojoudokan,I., Rolfe,M.,
Ruiz,S., Severy,G., Scherer,S., Scott,G., Shen,H., Shim,C.,
Shooshitari,N., Sisson,I., Sodergren,E., Sonaike,T., Sparks,A.,
Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabors,P., Tameria,A.,
Tameria,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B.,
Thomas,N., Thomas,S., Usmani,K., Vazquez,L., Vera,V., Villalon,D.,
Vinson,R., Wall,R., Wang,S., Ward-Moore,S., Warren,R.,
Washington,C., Watlington,S., Williams,G., Williamson,A.,
Wleszyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J.,
Zorrilla,S., Kucherlapati,R. and Gibbs,R.
Direct Submission

TITLE

JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 165236)
 AUTHORS Worley, K.C.
 TITLE Direct Submission
 JOURNAL Submitted (28-SEP-2000) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
 REFERENCE 3 (bases 1 to 165236)
 AUTHORS Worley, K.C.
 TITLE Direct Submission
 JOURNAL Submitted (30-NOV-2000) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
 COMMENT On Nov 30, 2000 this sequence version replaced gi:1125287.
 INFORMATION: <http://www.hgsc.bcm.tmc.edu/> or email gc-help@bcm.tmc.edu

CLONE LENGTH: This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the Features listing.

ANNOTATION OF FEATURES:

STSs are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.

Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished.) for Human and Mouse sequences.

Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL: <http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html>.

QUALSTAT-REPORT-----

```
----- Summary Statistics -----
Contig length:          165236
Phrap values in estimate: 164180
Fraction of Phrap values less than 40 : 0.000121131
Number of consensus changing edits: 0
Number of N's in consensus : 26
```

```
----- Consensus changing edits -----
Position      Original+Context      Edited+Context
4248          gtgcaggcac (n)acaaaaaaag  gtgcaggcac (a)acaaaaaaag
25468         tttaaagtt (n)atctctctt  tttaaagtt (t)atctctctt
35515         tgcctcagc (n)ccaattctt  tgcctcagc (t)ccaattctt
35635         acatctgaa (n)lactgggac  acatctgaa (c)lactgggac
36329         cagccctggc (n)acatgtgaa  cagccctggc (a)acatgtgaa
42433         ttctgttgg (n)tttaaaaca  ttctgttgg (t)tttaaaaca
52516         ttaattcag (n)catcatgtt  ttaattcag (t)catcatgtt
59538         caaagcaagt (n)acatggcaact  caaagcaagt (t)acatggcaact
```

```
90839      actgatcac (n)nganaagtat  actgatcac (a)aganaagtat
90840      ctgatcacn (n)ganaagtatg  ctgatcacn (a)ganaagtatg
90843      gatcacnnga (n)agaatcgtga  gatcacnnga (g)agaatcgtga
91026      caacactttg (c)gggacatttc  caacactttg (t)gggacatttc
101466      catcaatttg (n)tcacaacata  catcaatttg (t)tcacaacata
101608      aaataaacac (n)ccgttatgat  aaataaacac (c)ccgttatgat
109374      cctccacctt (t)cgataggccc  cctccacctt (c)cgataggccc
109693      accttatcgg (n)agaatgatct  accttatcgg (t)agaatgatct
109738      ctgggtcaaa (n)ggtatctctg  ctgggtcaaa (t)ggtatctctg
118157      aatcttaaaa (n)tcatgaacat  aatcttaaaa (t)tcatgaacat
139327      ggaacttagt (n)aanmatctt  ggaacttagt (g)aanmatctt
139327      tcttagtgaa (n)matcttctt  tcttagtgaa (t)matcttctt
139328      cttagtgaan (n)natcttctt  cttagtgaan (t)natcttctt
139329      ttagtgaann (n)attcttctt  ttagtgaann (t)attcttctt
139330      tagtgaann (a)tttcttctt  tagtgaann (t)tttcttctt
146052      ccttaaac (t)acaaaggcct  ccttaaac (g)acaaaggcct
146207      gtcatcagg (n)aatggaatt  gtcatcagg (a)aatggaatt
160897      accctcgag (n)tttagagta  accctcgag (c)tttagagta
```

----- Distribution of Quality < 40 Bases -----

```

1000
900
800
700
600
500
400
300
200
100
0
#
bases
* * * * *
5 10 15 20 25 30 35 40
Phrap Value Range
```

```
-----
Version: 1.01 gxf0.
FEATURES
source      Location/Qualifiers
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="12"
            /clone="RP11-31A23"
repeat_region 2..309
            /rpt_family="L1P"
            /rpt_family="L1P"
            complement (312..1362)
repeat_region /rpt_family="L1M4"
            complement (1368..1821)
repeat_region /rpt_family="L1P4S"
            /rpt_family="L1P2"
repeat_region 1822..2720
            /rpt_family="L1P2"
repeat_region 2705..5433
            /rpt_family="L1P1A12"
repeat_region 5434..5788
            /rpt_family="THE1B"
repeat_region 5789..6973
            /rpt_family="L1P1A12"
repeat_region 6974..7278
            /rpt_family="A1U5p"
repeat_region 7279..7425
            /rpt_family="L1P1A12"
repeat_region 7428..7874
            /rpt_family="L1"
repeat_region 7876..7959
            /rpt_family="L1MCc"
repeat_region 7960..7979
            /rpt_family="TAA)n"
repeat_region 7980..8265
            /rpt_family="L1MCc"
            complement (8266..8568)
```


TITLE
JOURNAL
REFERENCE
AUTHORS

TITLE
JOURNAL
COMMENT

Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J., Laroque, K., Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Lieu, C., Liu, G., Locke, K., Macdonald, P., Margulis, N., McCarthy, M., McEwan, P., McGuck, A., McKernan, K., McPheeters, R., Meldrum, J., Meneus, L., Mihova, T., Miranda, C., Mienna, V., Morrow, J., Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, T.M., Oliver, U., Peterson, K., Pierre, N., Pisan, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D., Roy, A., Santos, R., Schauer, S., Severy, P., Spence, A., Talmas, J., Strange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Tassef, S., Theodore, J., Turrell, A., Travers, M., Triggillo, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zimmer, A. and Zody, M.

Direct Submission
Submitted (04-MAY-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 174253)

Birren, B., Linton, L., Nusbaum, C., Lander, E., Abrahams, H., Allen, N., Anderson, S., Baldwin, J., Barna, N., Baetsen, V., Beda, F., Boguslavsky, L., Boukhgalter, B., Brown, A., Burkett, G., Campopiano, A., Castle, A., Choquel, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P., Dearlano, K., Dewar, K., Diaz, J.S., Dodge, S., Domino, M., Doyle, M., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L., Grand-Pierre, N., Grant, G., Hagos, B., Heath, A., Horton, L., Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J., Laroque, K., Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Lieu, C., Liu, G., Locke, K., Macdonald, P., Marcuis, R., McCarthy, M., McEwan, P., McGuck, A., McKernan, K., McPheeters, R., Meldrum, J., Meneus, L., Mihova, T., Miranda, C., Mienna, V., Morrow, J., Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, T.M., Oliver, U., Peterson, K., Pierre, N., Pisan, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D., Roy, A., Santos, R., Schauer, S., Severy, P., Spence, A., Talmas, J., Strange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Tassef, S., Theodore, J., Turrell, A., Travers, M., Triggillo, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zimmer, A. and Zody, M.

Direct Submission
Submitted (24-AUG-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jun 9, 2000 this sequence version replaced gi:17705204.
All repeats were identified using RepeatMasker:
Salt, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information
Center project name: L9496
Center clone name: 131 K16

----- Summary Statistics
Sequencing vector: M13; M7815; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 156518 bases at least Q40
Consensus quality: 155902 bases at least Q30
Consensus quality: 169236 bases at least Q20
Insert size: 170000; agarose-fp
Insert size: 171353; sum-of-contigs
Quality coverage: 3.5 in Q20 bases; agarose-fp
Quality coverage: 3.5 in Q20 bases; sum-of-contigs

----- NOTE: This is a 'working draft' sequence. It currently consists of 30 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequences as soon as it is available and the accession number will be preserved.

FEATURES	source
1	1232: contig of 1232 bp in length
1233	1332: gap of 100 bp
1333	3334: contig of 2002 bp in length
3335	3434: gap of 100 bp
3435	5627: contig of 2133 bp in length
5628	5727: gap of 100 bp
5728	7753: contig of 2026 bp in length
7754	7853: gap of 100 bp
7854	9492: contig of 1639 bp in length
9493	9592: gap of 100 bp
9593	11429: contig of 1837 bp in length
11430	11529: gap of 100 bp
11530	13178: contig of 1649 bp in length
13179	13278: gap of 100 bp
13279	15953: contig of 2675 bp in length
15954	16053: gap of 100 bp
16054	18335: contig of 2282 bp in length
18336	18435: gap of 100 bp
18436	19919: contig of 1484 bp in length
19920	20019: gap of 100 bp
20020	22351: contig of 2332 bp in length
22352	22451: gap of 100 bp
22452	25473: contig of 3022 bp in length
25474	25573: gap of 100 bp
25574	28192: contig of 2619 bp in length
28193	28292: gap of 100 bp
28293	32574: contig of 4282 bp in length
32575	32674: gap of 100 bp
32675	36983: contig of 4309 bp in length
36984	37083: gap of 100 bp
37084	41557: contig of 4474 bp in length
41558	41657: gap of 100 bp
41658	46775: contig of 5118 bp in length
46776	46875: gap of 100 bp
46876	52390: contig of 5515 bp in length
52391	52490: gap of 100 bp
52491	57638: contig of 5148 bp in length
57639	57738: gap of 100 bp
57739	65255: contig of 7517 bp in length
65256	65355: gap of 100 bp
65356	71907: contig of 6552 bp in length
71908	72007: gap of 100 bp
72008	79155: contig of 7148 bp in length
79156	79255: gap of 100 bp
79256	88577: contig of 9322 bp in length
88578	88677: gap of 100 bp
88678	98500: contig of 9823 bp in length
98501	98600: gap of 100 bp
98601	106888: contig of 8288 bp in length
106889	106988: gap of 100 bp
106989	117423: contig of 10435 bp in length
117424	117523: gap of 100 bp
117524	125991: contig of 8466 bp in length
125992	126091: gap of 100 bp
126092	137392: contig of 11301 bp in length
137393	137492: gap of 100 bp
137493	154901: contig of 17409 bp in length
154902	155001: gap of 100 bp
155002	174253: contig of 19252 bp in length.

Location/Qualifiers

1. 174253

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="8"

/map="8"

/clone="RP11-131K16"

/clone_lib="RPCT-11 Human Male BAC"

1. 1232

/note="assembly_fragment"

1333. 3334

/note="assembly_fragment"

3435. 5627

/note="assembly_fragment"


```

misc_feature      5728..7753
                    /note="assembly_fragment"
misc_feature      7854..9492
                    /note="assembly_fragment"
misc_feature      9593..11429
                    /note="assembly_fragment"
misc_feature      11530..13178
                    /note="assembly_fragment"
misc_feature      13279..15953
                    /note="assembly_fragment"
misc_feature      16054..18335
                    /note="assembly_fragment"
misc_feature      18436..19919
                    /note="assembly_fragment"
                    clone_end:17
                    vector_side:right"
misc_feature      20020..22351
                    /note="assembly_fragment"
misc_feature      22452..25473
                    /note="assembly_fragment"
misc_feature      25574..28192
                    /note="assembly_fragment"
misc_feature      28293..32574
                    /note="assembly_fragment"
misc_feature      32675..36983
                    /note="assembly_fragment"
misc_feature      37084..41557
                    /note="assembly_fragment"
misc_feature      41658..46775
                    /note="assembly_fragment"
misc_feature      46876..52390
                    /note="assembly_fragment"

```

Query Match Best Local Similarity 4.3%; Score 46; DB 2; Length 174253;
Matches 118; Conservative 0; Mismatches 120; Indels 0; Gaps 0;

```

QY 482 TCGGATTGTATCAACAATTATCTCAATCATCATCATGCTGCGAGTGTATATCAAGCCCT 541
Db 59082 TTGGAAGTAAATCTTAAATAAATAATCATGTCATTCGACGACATTCATTTGATGATTT 59023
QY 542 CTGATCTTCCACATCTGCGACCAAGTCAACAAGATATCATCTCAAAAATATGAAACCA 601
Db 59022 CAGGACTCCCTTGTATGTGTGCAACCTAAAGAAGCACTGATGTATGATGAG 58963
QY 602 TTGTGATTCACAGTGTGCTCTCAAGAGAAATATATTAATCTGTGAAGCAAGSTA 661
Db 58962 GTTATATCTACAGTTGCCATCTATATTAAGCTGTGATTTATTTTAAAAACCTGTCA 58903
QY 662 AACCAATTTCAACTACTATGTGCGATATAGTGAAGATATGTGAATTAAGGATAT 719
Db 58902 AACCAATCTAATTAAGTGTGCTGACTTCCCATTTGAATAAGCACTTGAATTT 58845

```

```

RESULT 10
AC124649/1 187254 bp DNA linear HTG 06-AUG-2002
LOCUS Homo sapiens chromosome 8 clone RP11-2318 map 8, *** SEQUENCING IN
DEFINITION
AC124649 187254 bp DNA linear HTG 06-AUG-2002
PROGRESS *** 2 ordered pieces.
AC124649
AC124649.3 GI:22123145
VERSION HTG; HTGS_PHASE2; HTGS_FULFILL; HTGS_ACTIVEFIN.
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE
1 (bases 1 to 187254)
AUTHORS Birren,B., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 8, clone RP11-2318
JOURNAL Unpublished
2 (bases 1 to 187254)
AUTHORS Birren,B., Lincoln,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavsky,L.,
Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,

```

TITLE
JOURNAL
REFERENCE
AUTHORS

Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collymore, A.,
Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S.,
Faro, S., Ferreira, P., Fitzgerald, M., Fitzhugh, W., Gage, D.,
Galagan, J., Gardyna, S., Ginde, S., Gord, S., Goette, M., Graham, L.,
Grand-Pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I.,
Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Lacroque, K.,
Lamazzar, R., Lander, T., Lehotzky, J., Levine, R., Lindblad-Toh, K.,
Liu, G., Maclean, C., Macdonald, P., Major, J., Margulis, N.,
Mathews, C., McCarthy, M., McEwan, P., McKernan, K., Meldrum, J.,
Menus, L., Mihova, T., Mienga, V., Murphy, T., Naylor, J., Nguyen, C.,
Nicola, R., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P.,
O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N.,
Pollara, V., Raymond, C., Retta, R., Ribback, M., Riley, R., Rise, C.,
Rogov, P., Roman, J., Roselli, M., Roy, A., Santos, R., Schauer, S.,
Schupbach, R., Seaman, S., Severy, P., Spencer, R., Stange-Thomann, N.,
Stojanovic, N., Straus, K., Subramanian, A., Talamas, J., Testfaye, S.,
Theodore, J., Topham, K., Travers, M., Travis, N., Triggillo, J.,
Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J.,
Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE
JOURNAL
COMMENT

Submitted (15-JUN-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 187254)
Birren, B., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S.,
Barna, N., Bastien, V., Bloom, T., Boguslavsky, L., Boukhgalter, B.,
Camarata, J., Chang, J., Chazaro, B., Choepel, Y., Collymore, A.,
Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S.,
Faro, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J.,
Gardyna, S., Gord, S., Graham, L., Grand-Pierre, N., Hagos, B.,
Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A.,
Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K.,
Liu, G., Maclean, C., Macdonald, P., Major, J., Mathews, C.,
McCarthy, M., Meldrum, J., Menus, L., Mihova, T., Mienga, V.,
Murphy, T., Naylor, J., Nguyen, C., Nicola, R., Norbu, C., Norman, C.H.,
O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K.,
Phunkhang, P., Pierre, N., Raymond, C., Retta, R., Rise, C., Rogov, P.,
Roman, J., Roy, A., Schauer, S., Schupbach, R., Seaman, S., Severy, P.,
Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Talamas, J.,
Testfaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J.,
Zembek, L., Zimmer, A. and Zody, M.

Direct Submission
Submitted (06-AUG-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Aug 6, 2002 this sequence version replaced gi:21699345.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/BW/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L27405
Center clone name: 23_I_8

* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 64780: contig of 64780 bp in length
* 64781 64880: gap of 100 bp
* 64881 187254: contig of 122374 bp in length.
Location/Qualifiers
1..187254
/organism="Homo sapiens"
/db_xref="taxon:9606"

* consists of 5 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1
* 3390 3389: contig of 3389 bp in length
* 3490 3489: gap of unknown length
* 19780 19779: contig of 16290 bp in length
* 19880 39080: contig of 19201 bp in length
* 39081 39180: gap of unknown length
* 39181 67046: contig of 27866 bp in length
* 67047 67146: gap of unknown length
* 67147 114741: contig of 47595 bp in length.
Location/Qualifiers
1. 114741
/organism="Medicago truncatula"
/db_xref="taxon:3880"
/clone_11b="Medicago truncatula BAC library H1"
BASE COUNT 39037 a 19145 c 18697 g 37461 t 401 others
ORIGIN

Query Match 4.3%; Score 45.4; DB 2; Length 114741;
Best Local Similarity 50.2%; Pred. No. 13;
Matches 112; Conservative 0; Mismatches 111; Indels 0; Gaps 0;
QY 604 TTGATTCAACAGTTGTCCTTCACAAAGAAATATTAATCTGTAAGCAAGTAA 663
Db 12845 TGGAAATCAGCAAAATTTATTTGATATATGAGATGTTAAATCAAAAGCAT 12786
QY 664 CCAATTTCAACTACTATGTCGATATAGTGAAGATAGTGAATTAAGATATCATC 723
Db 12785 TGCCTCAAAAATTTAATTAATTAATCTATGTTACCTTAAGTGAATATATCAATTCTAA 12726
QY 724 AGAGATCAACATCAATGTTCAACTCATTCCTCTGATTAAGCCATACGTTATACT 783
Db 12725 TAACATCTTCAACATGATTTCTGATTCAGTTATTTTATTAAGAAATCGGATTATGCT 12666
QY 784 GATGTTTGGCTTGAATTCGAATTTGTGATGAGAAACCAA 826
Db 12665 GTTGATTTTGTGTTGCTTTTGAATTCGATGAGTAAATGTCATA 12623
RESULT 13
R 13
L 13
DEFINITION REXX01 282610 bp DNA linear BCT 11-NOV-1998
Rickettsia prowazekii strain Madrid E, complete genome; segment
1/4.
ACCESSION AJ235270 AJ235269
VERSION AJ235270.1 GI:3860572
KEYWORDS complete genome.
SOURCE Rickettsia prowazekii.
ORGANISM Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
Rickettsiaceae; Rickettsiae; Rickettsia; typhus group.
REFERENCE 1 (bases 1 to 282610)
Andersson, S.G., Zomorodipour, A., Andersson, J.O.,
Scheritz-Ponten, T., Almark, U.C., Podowski, R.M., Naslund, A.K.,
Eriksson, A.S., Winkler, H.H. and Kurland, C.G.
The genome sequence of Rickettsia prowazekii and the origin of
mitochondria
JOURNAL Nature 396 (6707), 133-140 (1998)
MEDLINE 99039499
PUBMED 9823893
REFERENCE 2 (bases 1 to 282610)
Andersson, S.G.E.
AUTHORS Direct Submission
TITLE Submitted (11-NOV-1998) S.G.E. Andersson,
JOURNAL Stiv.Andersson@molbio.uu.se, Dept. of Molecular Biology, University

FEATURES
source of Uppsala, Husargatan 3, Uppsala, S-751 24, SWEDEN
Location/Qualifiers
1. 282610
/organism="Rickettsia prowazekii"
/strain="Madrid E"
/db_xref="taxon:782"

gene
CDS
1. 822
/gene="RP001"
/gene="RP001"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="CA14474.1"
/db_xref="GI:3860573"
/translation="MTKLIIHLVSDSSVOTAKHAANSALAOFTSIKQKLYHWPIMRC
ELLNEVLAKIESKHGIVLTYIADQELRKLTYCEYELKPCISIVGKI KENSVSFGI
EIAEKQNVYKFDKTYFDTLNADVAIRHDDCOMINLESSEDIILGPERTSGTSFV
FLAYNGLRKANIPYVNCPPPDPIEKDIDQVLVGLVFNRRLEIFEARLNLIQJEN
KSYTDPNIVORECIEVRKICNQRNMFVIDVSTRSLEETALIMRIYNRKNTYK"
906. 1298
/gene="RP002"
/gene="RP002"
906. 1298
/gene="RP002"
/codon_start=1
/transl_table=11
/product="THIOREDOXIN (trxA)"
/protein_id="CA14475.1"
/db_xref="GI:3860574"
/translation="MSCVNEITITLFEFSDNDINTTORIMVNNVYDSSFKNELIESDL
PVMVDPMAMGCPCKRLPIPIDEISKELQDKYVKLMNIDENPKTSEYGISIPFIM
LFRNGEQDKTKIGLQCKNSLDWINKSI"
1309. 2067
/gene="RP003"
1309. 2067
/gene="RP003"
/codon_start=1
/transl_table=11
/product="O-ANTIGEN EXPORT SYSTEM ATP-BINDING PROTEIN RPEB
rfbB"
/protein_id="CA14476.1"
/db_xref="GI:3860575"
/translation="MLNHSKVFIETDGVSGIDVHKAQGLKHFPLKKGVSLPTI
PLNNINFSCEYGEKIAFIGSNGSGSKLKLAIPIPKSGIVKHGIIAIIIDGV
GGEQGTGRKENIKMLVNNMLDKYSKKEIKIIFSELSGSKIDPIKYSQGMISRL
AESVSIPONPQIILLDEVPAGDSYFIKSLMLMKKPFNPTISIIIVSHQEBIIKQNC
DRCILKDGHIIDGTPSEIFKIYKQSNKEIHK"
2064. 2840
/gene="RP004"
2064. 2840
/gene="RP004"
/codon_start=1
/transl_table=11
/product="O-ANTIGEN EXPORT SYSTEM PERMEASE PROTEIN RPEB
rfbA"
/protein_id="CA14477.1"
/db_xref="GI:3860576"
/translation="MIKYFSSKKYWRALALVKAIVSRONKDSFLGSLMSLIQPIYHI
IYISYVFGFLRQPREFMVNLVGGIPIKSFVSLTICASSIVTRDQIIKKVRISKY
FPFIASLSQVLTILCSFSANVFATILPEKFSNQTIMPTLIIPIVIVISGSIAY
AFETPIRIRIQQMLNVLGVIIYSIPIVPISLIPSKKITYEENFFVIYIRVQVLV
INGSLPDMWLIIRKSVIVAFITVSIYLIYRFSKRVIIYL"
2837. 3724
/gene="RP005"
2837. 3724
/gene="RP005"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="CA14478.1"
/db_xref="GI:3860577"
/translation="MKLGFNLDPNEDLTGLKLDQIPLDIYLPKADSKLHKDMLMFLRA
TPLSIIPQDSKFLKISPHLDPLAELFCISKVEVTSRLKHKDPDIYECRRKQVOR

```

VAVKKYPLEKIDIDFEDVYLTITNLIGNFTSREFAKQIVIMQODESPSLDIAA
RYAAYVFSYVNSLSMNKQLFQNDILFNLOQKLDKQLIDKILKQKERVDFD
YDSFENLDEALNNSHCYCHKRDKDSCKGPDVAPHRDVAISVDKIPRLSYRMTGT
CPLKOKISEMYTIKAQVLI"
gene
3757. .5877
/ gene="RP006"
3757. .5877
/ gene="RP006"
/codon_start=1
/ transl_table=11
/product="unknown"
/protein_id="CA14479.1"
/db_xref="GI:3860578"
/translation="MVAATGHRICNDSCAKCIYQKODPVNIPLESNILEETKLTPYG
LEIYILLTRNPLNIYAPLPKEPTNIIIVTQSPAGFSLSYLILSGNVAIDLX
ITPLSNHKKPVKFMHEVKNLISERIPKRGVAYEYITIRMDKXNDLIRLILEBN
NKRYDVGALGRNITKEQALDPSFDVAPCTGAKPKIITNTEPEAKGRTASDPLT
LQNGAFVLONSNTMMIRMPVYIIGGLTSLDVATSLPYKQVESEKANYIEKDLT
BEDKEIAEEFIHAQKLFEKAKNNEELKVFNKLGATVYVRGLQSPAYKINHELI
YTLALGVNEKNNQPLRINIDKYHVESVEFTITLDRTHKKYHINMCPYKPSN
TVLTKTQVIMAGIENTOFDYKSYFGDCPKYEGSVAKITSAREGEVINKRL
INVPSEKSYACFITQDLYLTSRINKNINILNDKTEPILHSPLAKRPOQPFRL
QNSKDTKLIIEVALSPADIDIEKGLISFYVEGKSTSLCTLSSEKVVLMGFTG
SPLEIPQNKIIIDSKRYNVLKILKENKKNKVIATYPDIKKHULTSVDIVIINTS
PEIAELQELIKIFGENTELIIVNSLMQCMKGIQQCIQYKQKXIFACEQON
VEIDPKSLKTRLRONSLOEKSN"
complement (6227. .7021)
/ gene="RP007"
complement (6227. .7021)
/ gene="RP007"
/codon_start=1
/ transl_table=11
/product="ACYL- (ACYL-CARRIER-PROTEIN) --UDP-N-
ACETYLGUCOSAMINE O-ACULTRANSFERASE (lpxa)"
/protein_id="CA14480.1"
/db_xref="GI:3860579"
/translation="MSNSNIHTTAIIAEGAKFGKNGVPCYCIIGPEVILHNVELKS
GVVIDITEIGENTVIYPPASIGOPOLKYANERSTIIGSNITREYVIVQAGSKS
HGMITRVGNLNFVNGVHIGDCKIGNLVFANVYVLAGHIVKQVDAIIIGISAVHGY
TRIGESMTIGGSPVADYIPROLVSSKAVLEGLNLIGNRKGFPKADSLRANAVE
ETFLGGENFVDRIKOVAEKYKNSIVTQIIDFLNDSRAFCIFK"
complement (7031. .7468)
/ gene="RP008"
complement (7031. .7468)
/ gene="RP008"
/codon_start=1
/ transl_table=11
/product=" (3R)-HYDROXYMYRISTOYL- [ACYL CARRIER PROTEIN]
DEHYDRATASE (fab2)"
/protein_id="CA14481.1"
/db_xref="GI:3860580"
/translation="MIAITEIMDLIPHRYPILLVDRVLKIDPKNSIIGIKVATNEP
QETGHPARPVMPGVLMVESMAQALAILVAKSLDSTKNEVFLMSIENTKPRITVPG
DTMHISVIDOQANVMKFSKYMWECCEIAESKFTAMIKDKM"
complement (7659. .8699)
/ gene="RP009"
complement (7659. .8699)
/ gene="RP009"
/codon_start=1
/ transl_table=11
/product="UDP-3-O- (3-HYDROXYMYRISTOYL) GLUCOSAMINE
N-ACYLTRANSFERASE (lpkd)"
/protein_id="CA14482.1"
/db_xref="GI:3860581"
/translation="MWSNPFYKNGRKLTAIVDFLHDIIEPTKIVEDIDIDIKILQ
EASPNLISFLSNPKYSEFLKTKAACIYPKNTBEVNQNTVLIHENSYPFVYSKID
FFYAPIKSYSTCKIMKSAIADSAITIGKNCYIGNVIEDVIGDNIISADIGVIRG
VNIKGKARIEOHVSINVAIIGDQVILVAKIGODPGRSTEGVHKIPIHIGIVIG
NVEIGSNTTIRGALOPTIIEICRIDNVVOGHKVKIGKSIITYAQGIAGSSAIG
KTCALGGQVIGAGHLNIGSTQVAAQGVAAQNTIEGKIVGGSFAVPIIMHROSITMK
QLVKTSNLSKTK"
complement (9267. .9339)
/ gene="crRNA Phe (GAA)"

```

```

crRNA complement (9267. .9339)
gene / gene="crRNA Phe (GAA)"
9661. .10686
/ gene="RP011"
9661. .10686
/ gene="RP011"
/codon_start=1
/ transl_table=11
/product="NIFR3-LIKE PROTEIN (nifR3)"
/protein_id="CA14483.1"
/db_xref="GI:3860582"
/translation="MLNTDIMLNFITIMIKIGINELSSNVILAPMSDVTLEFFKLV
KRFAGLVASEMIASRAMIMKRSQMRQCAIMHDDPTSAVQIACGPVIAEAKMN
EDMGAKIIDLNGCDPAKRVGVAGSALMRDQLAKIPEATVKAVKIPVYKRMIGW
DNTNAPKIIPLAKIANSGVOMTVHGRTRCOFSGSNAMDFIRTEKAVKIPVINGD
ITNPAKAEALORSGADGIMRGVYKPMILISOIAYLYKKEKPAISIAEDLIIT
KHYDAITDYGKSVGPPIARKHIIWYNSGLPESAEFRCTVNLKMPIAVKEIAEFYM
SVMDNKK"
10683. .11756
/ gene="RP012"
10683. .11756
/ gene="RP012"
/codon_start=1
/ transl_table=11
/product="unknown"
/protein_id="CA14484.1"
/db_xref="GI:3860583"
/translation="MKRILSFIFILFFNSSYAVKAAIITDYKPIELVTEENKIK
IAIRSYLKNKSYFVLVDPNSEFKTEIVQELVILPANKIKENLILKLNKSYIALN
KYNFIDQISQSFQKDFSNKPTDNTIYVEHKLNSKSSHSSTPMNSTVOONVATSMY
Query Match 4.2% Score 44.6; DB 1; Length 282610;
Best Local Similarity 47.9%; Pred. No. 16;
Matches 128; Conservative 0; Mismatches 139; Indels 0; Gaps 0;
QY 708 AATTAAGATATCATCAGAGATCAACATACATGTTCAACTCATTGTAACCTGTGATAA 767
|||||
DB 254837 AATTAAGATATCATATTAATTAATCAATAAATTAATCAATATATTAATTTTATA 254896
|||||
QY 768 GCATACGTTTATACGATGTTTGGCTTGAATTCATTTGTGATGAGAAAGCAAAA 827
|||||
DB 254897 GAAGAAATTTTGAATTTGAATTTTATTAATTAATTAATGCTTTTAAAGCAAAA 254956
|||||
QY 828 GGTGCTGTGAAGTATACAGAAATTTATGACTAATTAATGTTGACTTGTGG 887
|||||
DB 254957 CATATATCTTGTGTGAAGATTAATTTTATGATTAATTAATGATGCTTATTAATTC 255016
|||||
QY 888 TCTCGATTACTCTCCAGCCAAAGATGATGCTCATTTGGCTAAATCATCAAA 947
|||||
DB 255017 CATGGGTTAACCGTTTCAGTTCTTATTAAGGTATAGTGTATGCTATAGAGAGAAA 255076
|||||
QY 948 CTTTATGCTCAATTGAGCCACAAATT 974
|||||
DB 255077 ATTTGATTAAGAAATTAAGCTTAATAACT 255103
|||||
RESULT 14
AP004829/c 304050 bp DNA linear BCT 02-JUL-2002
LOCUS Staphylococcus aureus subsp. aureus MM2 DNA, complete genome,
DEFINITION strain:MM2, section 8/10.
ACCESSION AP004829 BA000033
VERSION AP004829.1 GI:21205117
KEYWORDS
SOURCE
ORGANISM Staphylococcus aureus subsp. aureus MM2 (strain:MM2) DNA.
Bacteria; Firmicutes; Bacillales; Staphylococcus.
REFERENCE
1 Baba, T., Takeuchi, F., Kuroda, M., Yuzawa, H., Aoki, K., Oguchi, A.,
Nagai, Y., Iwana, N., Asano, K., Nami, T., Kuroda, H., Cui, L.,
Yamamoto, K. and Hiramatsu, K.
Genome and virulence determinants of high virulence
community-acquired MRSA

```

JOURNAL Lancet 359 (9320), 1819-1827 (2002)
MEDLINE 22040717
PUBMED 12044378
2 (bases 1 to 304050)
AUTHORS Aoki, K., Oguchi, A., Nagai, Y., Asano, K., Iwama, N., Baba, T.,
Kuroda, M., Hiramatsu, K. and Kikuchi, H.
TITLE Direct Submission
SUBMITTED (06-MAR-2002) Aki Oguchi, National Institute of
Technology and Evaluation, Biotechnology Center; 2Chome 49-10
Nishihara, Shibuya-ku, Tokyo 151-0066, Japan
(E-mail:oguchi@nite.go.jp, URL:http://www.bio.nite.go.jp/
Tel:81-3-3481-8423, Fax:81-3-3481-8424)
FEATURES Location/Qualifiers
1. .304050
/organism="Staphylococcus aureus subsp. aureus MW2"
/strain="MW2"
/db_xref="taxon:196620"
/complement (629. .1072)
/gene="MW1947"
/complement (629. .1072)
/gene="MW1947"
/note="ORFID:MM1947
hypothetical protein, similar to acetyltransferase
[genomic island nu Sa 4mw]"
/codon_start=1
/transl_table=1
/protein_id="BAB95812.1"
/db_xref="GI:21205118"
/translation="MSITRLPNNSDPEKLNOCUKYDLDGYPPTNENDLKKRLKITT
HDDYFLILLIKENKIIGLSGCKMPEKRAEYMRILAIVHSEFRKKYGRKLLADS
EERSKLNCATILNSGNRNERLSAHLKYSIDNGVNTSGFTKOL"
/complement (1178. .1696)
/gene="MW1948"
/complement (1178. .1696)
/note="ORFID:MM1948
[genomic island nu Sa 4mw]"
/codon_start=1
/transl_table=1
/product="hypothetical perotein"
/protein_id="BAB95813.1"
/db_xref="GI:21205119"
/translation="MDNINKERKRVLSIAITISPLNGLISLAINGMKNKO
CISLNLSDTFAVISWTLTGALIGLAPSIYGFSPIDKSSFRASISWEGEO
RNWVKDMLMLPFLTSLILVIFIFQPAVSKLFTNTAIIIFIIILVMFPLILLICVK
ITWICRYFSKE"
/complement (1932. .2522)
/gene="MW1949"
/complement (1932. .2522)
/note="ORFID:MM1949
hypothetical protein, similar to bacteriophage terminase
small subunit [genomic island nu Sa 4mw]"
/codon_start=1
/transl_table=1
/protein_id="BAB95814.1"
/db_xref="GI:21205120"
/translation="MSELTAQAFVNEYIRTLNVTOSAIKAGYNANSAHTACRLLK
KPHIKOYIOKOKIIDENVLTAKELHLVLTNATVGBETKEVYVVRGGRKKNPQSG
KQVLVYEHVELIEVPKPSDRLLKARMLSKYHKLPLFDKIDINGNVPIFINIGEMDGG
DEGLDKAVKDVNSDNPHTIVDDIPLDYESTRFL"
/complement (2519. .2734)
/gene="MW1950"
/complement (2519. .2734)
/note="ORFID:MM1950
[genomic island nu Sa 4mw]"
/codon_start=1
/transl_table=1
/product="truncated hypothetical protein"
/protein_id="BAB95815.1"
/db_xref="GI:21205121"
/translation="MMTDDNVNDHIIKNHIEMLIVDLATDKEFYIPDSLIIQGSYODI

gene
2864. .3338
/gene="truncated int"
2864. .3211
/gene="truncated int"
/note="ORFID:MM1951
truncated hypothetical protein, similar to integrase
[genomic island nu Sa 4mw]"
/codon_start=1
/transl_table=1
/protein_id="BAB95816.1"
/db_xref="GI:21205122"
/translation="MKMSDIDFENNITSTTKTYVDPNNKKKFOILPKPTSSIGKIS
VDPNATKVRDYKIVQNNWKNELINDHCKNKTINVDINHTVTOHOKYVALIQT
HTLRIAHRICTH"
3072. .3338
/gene="truncated int"
/note="ORFID:MM1952
truncated hypothetical protein, similar to integrase
[genomic island nu Sa 4mw]"
/codon_start=1
/transl_table=1
/protein_id="BAB95817.1"
/db_xref="GI:21205123"
/translation="MTTMYKLSMWIQTINSGRTSLIKNISTHSPRYTHCALLIESGV
HIKEIOERLRKHDINTMYIAKITNSYKDKASHKSHOMEDVS"
/complement (3433. .5049)
/gene="groEL"
/complement (3433. .5049)
/note="ORFID:MM1953"
/codon_start=1
/transl_table=1
/product="GroEL protein"
/protein_id="BAB95818.1"
/db_xref="GI:21205124"
/translation="MYKQKPSFSDARQAMLRGVDOANAVKVTIGPKGNVLDKEFT
APLITNDGVTLKELIEDPYENMGAKLVQEVANKTNELAGCTTATYLAQMITEG
LNVISGANPVGLRQGIDKAVVAEALHENSQKKNKEIAQVIAISADEIGRYI
SEMERKVGNDGVITTEESNGNLTELEVGMQFDRGQSPYVWTSDDKVALEIRPYI
LVYDKKISFOOILPLLEQVOSNRPLIIVAEVEEDALTNVLNMRGTFPAVAKA
PGCDRRKAMLEDIALTGAOYITDGLIDGASIDMGTSKVEKASDETTLVVDG
GBENSIDARVSQKSQIETESDPPREKIQEHLAKLGVAVIKGASABTELKEKRL
RIEDANLSTRAAVEGIVAGGGLVNVQVSELEABSDITGVNVLKALTAVRQ
IAENGLTESVVERLKNAPGVPAATNEWMNMLEAGIVDPTKVRSALOHAASVA
AMFLTEAVEVASIPEKNDQPNMGWPGHM"
/complement (5125. .5409)
/gene="groES"
/complement (5125. .5409)
/note="ORFID:MM1954"
/codon_start=1
/transl_table=1
/product="GroES protein"
/protein_id="BAB95819.1"
/db_xref="GI:21205125"
/translation="MLKPIGNRVITTEKBOOTKSGIVLTSAKESNEGIVAVGT
GRILNIGTVTEPEVEDRIVFQQVAGTEVKSDNETYLVLINEEDIAVE"
5584. .6327
/gene="MW1955"
5584. .6327
/note="ORFID:MM1955"
/codon_start=1
/transl_table=1
/product="conserved hypothetical protein"
/protein_id="BAB95820.1"
/db_xref="GI:21205126"
/translation="MTRLWASLTLVLIYLSQFLPLIIVKLPVQVSGIEELTKAVIY
IOVLPLIATTIILINKIKNPTELEVKBPKYIIPALGFLVLMVQVAVSI
LVQIVGGQOVSPNTEKLIITARIKPIETFFVSIIGPLIEEVFRKVIPEGLPNAIKGN
RVAFIATIVSSLIIPALANDKPIPVYFGVGIVFSLAVWTKRLAVPIIIHMLONG
FVIVFQILNPEALKRATQANFIYHIFIP"

gene
CDS
/translatability="WTKDLILAETSCDETSVIRKNGRDIISNTLSIOESKPRG
GVPEVAVRHHCEGTTITINELALVDVSMEDIDAIATVBEPGLIGALLGVNAKRL
AFAYDKPLIPVHILIGHIVANHIEEPLEPLILAIIVSGGHEIVMKDHSFEVIGET
RDDAAGHGVKARITGLINPGQPVRLAEGEDTYSFPPEWLDKDSYDPSFGSLKS
AVINLHONROKNIPIIEANVATCFONSIVLVLFKAIQACKESYORLIYAGVASN
KGLRSLADQCKVMDIQGLTIPSPKCTDNAMMIGVAGHSIVQGRFADLNLNHSNID
LEEYSAP"
complement (3240. . 3704)
/gene="SA1855"
complement (3240. . 3704)
/gene="SA1855"
/note="ORFID:SA1855
hypochemical protein, similar to ribosomal-protein-alanine
N-acetyltransferase"
/codon_start=1
/transl_table=1
/protein_id="BAB43137.1"
/db_xref="GI:13701845"
/translation="WDQSKEDLNIREMTKEDVPQVPLERSFNDSWTIDAFHEIT
EONNPAKFLVEFEQQLIGYGLMIVIDQAITVALIDNTRGYGLQMLTKGINTYA
SHTCMVSLERVANNKVAQHYENLGFOYGGKRNXYGEGEDAMVWVNLND"
complement (3677. . 4366)
/gene="SA1856"
complement (3677. . 4366)
/gene="SA1856"
/note="ORFID:SA1856
hypochemical protein, similar to glycoprotein
endopeptidase"
/codon_start=1
/transl_table=1
/protein_id="BAB43138.1"
/db_xref="GI:13701846"
/translation="MNKLRRQLSNMSLLIDTSNOPLSVALMONDEVIAEITTSKQNH
SVQMLPALSQLEQSGIAKQOQDAIIVAEQSGYGLRIGTVAKTILVALDVLYGV
SLKLKALTIHDTKDLIVPVPARQAVYTGIFQWNGQLVTILEDQYMSIALRAFEL
EDLNPFYICGDTTKIDQDELQGEVIAQDPNASVWYQILDRPTVHTFTPKYHLAAR
ENMNSIQKN"
complement (4320. . 4781)
/gene="SA1857"
complement (4320. . 4781)
/gene="SA1857"
/note="ORFID:SA1857"
/codon_start=1
/transl_table=1
/product="conserved hypochemical protein"
/protein_id="BAB43139.1"
/db_xref="GI:13701847"
/translation="WIKINLDEMNQPMFLVEQKSGDILLNGDLAGKTTTQRI
GALVGRRIINSPFNIIKSYRKNLKLHMDCYRLDESDDLGEDEFEEQALTVIE
WSQFKDLIPATHLSINISTISENTRQIETLPAQGEHYEQIKEAIHHEFAH"
5292. . 6980
/gene="11VD"
5292. . 6980
/gene="11VD"
/note="ORFID:SA1858"
/codon_start=1
/transl_table=1
/product="dihydroxy-acid dehydratase"
/protein_id="BAB43140.1"
/db_xref="GI:13701848"
/translation="WRSDMIKKGDQAPARSLLAHTGALSKPTDMNKPFVAICNSYID
IPEVGHVHLEADIKAKEAREGAIPEEFNTIGVDDGLAMGHIKMRYSLEPEREIIADP
AEVIVANHEFDGVFIYPNCDKITPGMIIFAANTVNDPAIFCGGPKKALSHGKALLTV
SMPEFVAVKREKGSKEEFLDMEDNACPTGQSCGMETASMNCIMTEVIGLALPVNG
TLAVSDQREKIRQAPKLVENIKNDKRPDIYVREAIIDAPALDMMAGSSTVTVLH
TLIAINEADIIVDLERINAIARTPYLSKIAVSSYSMAHDVHAAGVPAIINELMKDK
GLHPIRIIVTGTITRENNEGKEIKNPVHILPDLAPYDAQGLSLFGNIIKPKAVIK
VGVGDPSTITFGKAIICFNSHDEAVEAIDNRTVRAGHVIVIREBPKGPPMPLMAR
TTSIYGRGSGKQVALITDGRFSGATRGIAVGHISPEAASGGPIALIEDGDEITDNLN
RTIANQGPEDVILARRRESLITPPKAVKTYGLARTALVTSANTGVMQVPENL"
7008. . 8777
/gene="11VB"
/gene="11VB"
7008. . 8777

```

/gene="l1aB"
/note="ORFID:SA1859"
/codon_start=1
/transl_table=11
/product="acetylactate synthase large subunit"
/protein_id="BA93141.1"
/db_xref="GI:13701849"
/translation="MSKTOHEVNQNIIDPLKMAEFLLEJOINLETLNDMRSGELVEA
LTLENVDYLFGRPGAVLPDYPTFYDGKIKHILARRHQGVAAEGYARSGKTGVV
VTGSPGATNWTGITDAHCDSLPVVFITGVVAITFGIGKDFQEDILSTSPITKQNY
QVRAREVIDPKRIAEAFHNVASGRKGVPVJDSFPMDGLVATNVDLCOEINPGEVATE
PENKDIDTIFSLKEAKKPVLVAGAGINSKSQMLTFQPNRKOIETVTLTLDLGAIVP
YEOTLFIQMGMGHGSYAASNALTECOLNLGSRFPDRLASKDPADAPNAKIIHVHID
PSKINLVYHVDLGIITADCKRFLECLNDKHVETIESHDPMWYGHCNNOKHPFKXGEEO
VEPKPOOTIFYIKTINGEAIVTTTDVGQHMAAOCPYPFNHGOWTSSGLGTMGEGEI
PSSIIGKLAMPDKTVCFVGDGGFQMTNQMALPEGLDYKIVLINNETLGHWKQMO
DKFENGRFSHSVFNGQPDFEMKAAYGVNGFLLIDKPLEQELEDAAFAVQGPVLEVR
ISPTLEATPMVPGSKNHMEGL"
8777 . 9031
/gene="SA1860"
8777 . 9031
/note="SA1860"
/note="ORFID:SA1860
hypothetical protein, similar to acetylactate synthase
small subunit"
/codon_start=1
/transl_table=11
/protein_id="BAB3142.1"
/db_xref="GI:13701850"
/translation="MTRILKLQVADVSTLNRTSAPVRLQVNIIDLTVHTSEOPGIS
NMEIOVDIODTSLHLIKLKQOINVLFVECYDLVDNEA"
9168 . 10172
/gene="l1cC"
9168 . 10172
/note="ORFID:SA1861"
/codon_start=1
/transl_table=11
/product="alpha-keto-beta-hydroxy-lactil reductoisomerase"
/protein_id="BAB3143.1"
/db_xref="GI:13701851"
/translation="MTTVYYDDVDVKTDALQGKKIAVVGYSQGHAAHQNKDGYDVV
IGIRGPSDFRKAGEDGVPPVAEAVQDAVINVLVLPDEIQGVYNKEIPNLEKNVA
LAHFAGHNHFRGVLOPADVDPLVAPKGGHVVRRFVGSVSPSLFGIOODASCQA
RNIALSYAKGICATRGAVIETTKEELENDLPEGOVLVCGSGKLIOSGETLEVAGY
OPELAYREULHEMKLIYDLMYBEGMEMVRYSISNTAEFCDPYVSGPRITTPDVKENMKA
VLDIIONGNFSNRPIEDNKNGFKEFYKLREEQHQLIEKVREUREMPITKSKSIEK
"
10202 . 11731
/gene="l1eua"
10202 . 11731
/note="l1eua"
/note="ORFID:SA1862"
/codon_start=1
/transl_table=11
/product="2-isopropylmalate synthase"

Query Match      4.2%; Score 44.6; DB 1; Length 307750;
Best Local Similarity 44.7%; Pred. No. 15;
Matches 173; Conservative 0; Mismatches 214; Indels 0; Gaps 0;

Cy      635 AATATATTACTCTGTAAGAACAAGTAACCAATTTCCAACTACTATGTCGATATAGTG 694
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 256627 AAAGTAGTAAGCTTGAATGTTTGTGTAATATGTAAAAACAACATATGATATATATGC 256568
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Cy      695 AAAGTAGTGTGAATTAAAGCATATCATCAGAGATCAACAATACATGTTCAACTCATTTG 754
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 256567 AATGATGTAGCTGATGATGAAGAAGTTGATGATGATACTCATTCATTTAAAGATAATG 256508
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Cy      755 GTACCTCTGATTAAGCCATAGCTTTATACTGATGTTTTGGCTTTGAATTTCCAATTTTGCTG 814
          ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 256507 GCATGCCGTTAATGATTTAAAGCTCTCTCTTTTATATTTTAAATTTTGTGACCAAGCC 256448
```


QY 815 ATGAAAAGCAAAAGCTTCTGTTGAGTTATCAAGAATTATTGACTAATCTTTAGTT 874
Db 256447 ATCCAATTAAATATMAAAATATCAAAACACGTTGATATAGATTAAATTTTAAAG 256388
QY 875 TGGACTTGTGGGCTCGGATTAACCTCCAGCCACAAGAATGATATGCTCATTTGG 934
Db 256387 TTTACATATACAGTTAAGATACAAAATATTCAGACTAAATTTCAATTTGGCAAAATAT 256328
QY 935 CTAAATCATCAAACTTTATGCTCAATTTGAGCCACAATTCGATGCCAAGAAAGTGAAG 994
Db 256327 CTTAAACATCAAAATTTATTAAGAAATTAATGTATTTAACCATATTTCTAGTAATAAAAAT 256268
QY 995 TTAGAGTTTGAAGTGTGACTTTG 1021
Db 256267 ATTGAATTTTATCTAATTTGTTAG 256241

Search completed: January 27, 2003, 05:30:06
Job time : 3762 secs

THIS PAGE LEFT BLANK

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 27, 2003, 03:13:59 ; Search time 221 Seconds
(without alignments)
10882.960 Million cell updates/sec

Title: US-09-675-509-3

Perfect score: 1068
Sequence: 1 atgtccaccacccaagaac.....gtgtcttgagaccattcctt 1068

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

1: N_Geneseq_101002.*
2: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
24: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
25: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	43.8	4.1	14041	22	AAH48024
2	43	4.0	3346	18	AAV74457
3	42.2	4.0	50000	24	ABL56201
4	41.6	3.9	3006	24	ABO69272
5	41.6	3.9	3024	24	ABO67866
6	41.6	3.9	40862	24	ABL34073
7	41.6	3.9	81905	24	ABO69244
8	41.6	3.9	82689	24	ABO67198
9	41.2	3.9	7491	21	AAV70156

10	41	3.8	9728	24	ABK40092	Human chemically p
11	41	3.8	9728	24	ABL33903	Human immune syste
12	41	3.8	15518	24	ABL70607	Chemically treated
13	41	3.8	15518	24	ABL34172	Human immune syste
14	41	3.8	15518	24	ABL34624	Human metastasis a
15	40.8	3.8	1537	22	AAFS8662	Porcine myoblastin
16	40.6	3.8	5407	22	ABL34091	Human immune syste
17	40.4	3.8	7969	22	AA546679	Tumour suppressor
18	40.4	3.8	7969	22	ABO67037	Human angiogenesis
19	40.2	3.8	1572	20	AA61708	B. burgdorferi ant
20	40.2	3.8	8245	22	AA546447	Tumour suppressor
21	40	3.7	4590	22	AAH24065	Yeast AOD9604-asso
22	40	3.7	17703	24	ABK39952	Human chemically p
23	39.8	3.7	6277	24	ABL33315	Human immune syste
24	39.8	3.7	13511	24	ABL32280	Human immune syste
25	39.6	3.7	2282	24	ABN95694	Gene #2192 used to
26	39.6	3.7	2287	16	AA700785	Human A6am1n CDNA.
27	39.6	3.7	2287	16	AAQ92617	Human A6am1n CDNA.
28	39.4	3.7	8842	24	ABK33449	Human immune syste
29	39.4	3.7	8842	24	AA563335	Chemically pretrea
30	39.4	3.7	19659	24	ABL32766	Human immune syste
31	39.2	3.7	113515	24	ABL34174	Human immune syste
32	39	3.7	2324	20	AAZ20302	Borrelia burgdorfe
33	39	3.7	6973	24	ABL33778	Human immune syste
34	39	3.7	15767	24	ABL33206	Human immune syste
35	39	3.7	15767	24	ABL34552	Human metacasis a
36	39	3.7	580073	18	AA758840	Mycoplasma genital
37	38.8	3.6	5371	24	ABL33187	Human immune syste
38	38.8	3.6	6106	22	AA546429	Tumour suppressor
39	38.8	3.6	6106	24	ABK40031	Human chemically p
40	38.8	3.6	6106	24	ABL33472	Human immune syste
41	38.8	3.6	7011	24	ABK39940	Human chemically p
42	38.8	3.6	7011	24	ABL32546	Human immune syste
43	38.8	3.6	7067	24	AA595269	Long terminal repe
44	38.8	3.6	17967	24	ABL33015	Human immune syste
45	38.6	3.6	3923	22	ABA08191	Human ovarian and

ALIGNMENTS

RESULT 1
AAH48024
ID AAH48024 standard; DNA; 14041 BP.
XX
AC AAH48024;
XX
DT 18-SEP-2001 (first entry)
XX
DE Internal control B19c #1.
XX
KW Internal control; ss.
XX
OS Parvovirus.
XX
PN WO200146463-A2.
XX
PD 28-JUN-2001.
XX
PF 20-DEC-2000; 2000MO-EP12996.
XX
PR 22-DEC-1999; 99AT-0002170.
XX
PA (BAXT) BAXTER AG.
XX
PI Zimmermann K, Turecek P, Schwarz H, Rieger M;
XX WPI; 2001-408656/43.
XX
PT Internal standards useful for nucleic acid amplification assays,
PT comprises a synthetic nucleic acid made by non-recombinant techniques
PT
XX

QY 490 TATCAATATCTTCCATATCATCATCAGCTGATGATATCAAGCCTTGATCTT 549
DB 2855 ATTATCTATCATATATAGATATATCATTAATANNNNNNNNNNNNNNNNNNNNNN 2796
QY 550 CCACATCTGTGTGACCAAGTCAACAGATATCTCAAAAATATAGAACCTTTGGAT 609
DB 2795 NNN 2736
QY 610 TCAACAGTGTGGCTCTCAAGAGATATATATATCTGTGAAGCAAGTAAACCAATT 669
DB 2735 TCAACGCGACATGACCTTATATGCTAAATGTTAACATGACTATATGTGATCTACT 2676
QY 670 TCAAACTACTATGTGGATATAGTAAAGTATGTGAAATTAAGATATCATCAGAGAT 729
DB 2675 TAAGATTGATATTTATGATTAAGATGATTTTAAGTTTAACTCA-ATATCCACATC 2618
QY 730 CAACATATCAATGTTCAACTGATGTTACTCTGATTAAGCC 770
DB 2617 TATTAATAATATGTTTAACTGATTAAGTCTTATACACC 2577
RESULT 3
ABLS6201
ID ABL56201 standard; DNA; 50000 BP.
XX
AC ABL56201;
XX
DT 01-JUL-2002 (first entry)
XX
DE AMEPV genome fragment#3.
XX
KM AMEPV; gene therapy; viral vector; chromosome mapping; gene mapping;
XX
KM genetic deficiency disorder; ds.
XX
OS Amsacta moorei entomopoxvirus.
XX
PN WO200212526-A2.
XX
PD 14-FEB-2002.
XX
PF 10-AUG-2001; 2001WO-US25287.
XX
PR 10-AUG-2000; 2000US-224479P.
XX
PR 14-SEP-2000; 2000US-0662254.
XX
PA (UYEL) UNIV FLORIDA.
XX
XX Moyer RW, Li Y, Bawden AL;
XX
DR MPI; 2002-227161/28.
XX
XX
PT Novel recombinant entomopox virus vector useful for delivering
PT polynucleotide encoding protein to vertebrate cell, comprises
PT polynucleotide encoding protein operably linked with heterologous
PT promoter sequence -
XX
PS Disclosure; Page 175-201; 326pp; English.
XX
XX The invention relates to a recombinant entomopox virus (EPV) vector,
CC comprising a polynucleotide encoding a protein operably linked with a
CC heterologous promoter sequence. The invention also concerns methods for
CC providing gene therapy for genetic deficiency disorders. Vectors of the
CC invention are useful for delivering a polynucleotide encoding a protein
CC to a vertebrate cell preferably a mammalian cell, such as a human cell.
CC The vector is introduced into the vertebrate cell by infection in a viral
CC or in vivo. The vector is useful for the delivery and expression of
CC biologically useful proteins in gene therapy protocols, and for
CC delivering large DNA segments for engineering of vertebrate cells.
CC Polynucleotides of the invention have applications in techniques such as
CC their use as insertion sites for foreign genes of interest, hybridisation
CC probes, for chromosome and gene mapping, in PCR technologies, and in the

CC production of sense or antisense nucleic acids. Vectors of the invention
CC provide for stable integration and expression of heterologous DNA in host
CC cells, and are adapted for accepting large heterologous polynucleotide
CC inserts which can be delivered in an infected or transformed cell and
CC expressed in a stable fraction. The current sequence represents a
CC fragment of the genome of the genus B entomopoxvirus from amsacta moorei
CC (AMEPV).
XX
SQ Sequence 50000 BP; 19754 A; 4704 C; 4365 G; 21177 T; 0 other;
Query Match 4.0%; Score 42.2; DB 24; Length 50000;
Best Local Similarity 46.2%; Pred. No. 2.5;
Matches 140; Conservative 0; Mismatches 163; Indels 0; Gaps 0;
QY 566 AAGTCACAGAGATATATCATCAAAATATAGAACATTTTGGATTCACAGTTGGCT 625
DB 33003 AATTATATCATTTTATATCATTTTAAATATATATCTTCTGATTTTGGATTTGATTTGCT 33062
QY 626 CTCAAGAGATATATATATCTGTGAAGCAAGTAAACCAATTTCAACTATGCTG 685
DB 33063 ATTATTAATAATATATATATTAATTAATTAATTAATTAATTAATTAATTAATTA 33122
QY 686 GATATAGTGAAGTATGTGGAATTAAGATATCATCAGAGATCAACATATACATGTT 745
DB 33123 GCAATTAT 33182
QY 746 AACTGATGTTGACCTGATAGCAATAGCTTATATCTGATTTTGGCTTGAATTC 805
DB 33183 ATTATATTAATTTATTTATGCAAAATATATATATATATATATATATATATATTT 33242
QY 806 ATTTGTGTGATGAAGCAAGAGTTGCTGTTGAAGTTATCAAGAAATTTATGCTAATA 865
DB 33243 TAATPACTTTTATTAATAATATATATATATATATATATATATATATATATATAT 33302
QY 866 CTT 868
DB 33303 ATT 33305
RESULT 4
ABQ69272
ID ABQ69272 standard; DNA; 3006 BP.
XX
AC ABQ69272;
XX
DT 29-AUG-2002 (first entry)
XX
XX
DE Listeria innocua DNA sequence #711.
XX
XX Antbacterial; Listeria; food contamination; mutational analysis;
XX infection; ds.
XX
XX Listeria innocua.
XX
PN WO200228891-A2.
XX
PD 11-APR-2002.
XX
PF 04-OCT-2001; 2001WO-FR03061.
XX
PR 04-OCT-2000; 2000FR-0012697.
XX
PA (INSP) INSR PASTEUR.
XX
PA (CNRS) CNRS CENT NAT RECH SCI.
XX
PI Kunst F, Glaser P;
XX
DR MPI; 2002-332479/37.
XX
XX New genomic sequences from Listeria species, useful for detection,
PT treatment and prevention of infection, also related polypeptides,
PT antibodies and modulators -
XX

PS Claim 7; SEQ ID 2085; 180bp; French.

XX The present invention relates to nucleic acid sequences
CC (AB067188-AB071212) from *Listeria* sp. The sequences are useful as probes
CC and primers for identification and/or detection of *Listeria* (e.g. as
CC contaminants in foods, or mutational analysis) and for analysis of
CC gene expression. Proteins encoded by the nucleic acid sequences can be
CC used to screen for compounds that modulate gene expression, replication
CC and pathogenicity of *Listeria* (potential therapeutic agents), also for
CC treating infections by *Listeria*, and are useful as immunogens in
CC anti-*Listeria* vaccines.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic format
CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 3006 BP; 1108 A; 406 C; 581 G; 911 T; 0 other;

XX Query Match 3.9%; Score 41.6; DB 24; Length 3006;

XX Best Local Similarity 50.5%; Pred. No. 1.4;

XX Matches 101; Conservative 0; Mismatches 99; Indels 0; Gaps 0;

XX 43 TATCTTCCTTCTTGGAATGGAATGGCAAGATTAAATGATCAATTGATCAAGAT 102

XX 475 TATATATTTTCAAAATTTGAAATGAAATATACACACATTTCTTAATGCGTTGAGCAA 534

XX 103 GTTTGGCAACTCAGGTTCCGATATATATCGAATATACCGAATTTGATTGTTACAGT 162

XX 535 GTTGGCATATACAGATTTATGTCAGAGTCAAGATACAGATTTAGATATATTACG 594

XX 163 GATGCTAGTCTTCAAGTCTTCCAGATGTTTCTCAACTGATGATTTCTTCATAT 222

XX 595 GCTGCATATCAAGAAATGCCCCAGAGTTATCTATATTTGCGTTATATATTGTTT 654

XX 223 CTGTTCTCTTGGGATGAT 242

XX 655 GGTGAATTTTGTCTGATTT 674

XX Db

XX RESULT 5

XX AB067866

XX ID AB067866 standard; DNA; 3024 BP.

XX AC AB067866;

XX DT 29-AUG-2002 (first entry)

XX DE *Listeria innocua* DNA sequence #668.

XX Antibiobacterial; *Listeria*; food contamination; mutational analysis;

XX infection; ds.

XX OS *Listeria innocua*.

XX PN WO200228891-A2.

XX PD 11-APR-2002.

XX PF 04-OCT-2001; 2001WO-FR03061.

XX PR 04-OCT-2000; 2000FR-0012697.

XX PA (INSP) INST PASTEUR.

XX PA (CNRS) CNRS CENT NAT RECH SCI.

XX PI Kunst F, Glaser P;

XX WPI; 2002-332479/37.

XX DR

XX PT New genomic sequences from *Listeria* species, useful for detection,

XX PT treatment and prevention of infection, also related polypeptides,

XX PT antibodies and modulators

XX PS Claim 7; SEQ ID 679; 180bp; French.

XX The present invention relates to nucleic acid sequences

CC (AB067188-AB071212) from *Listeria* sp. The sequences are useful as probes
CC and primers for identification and/or detection of *Listeria* (e.g. as
CC contaminants in foods, or mutational analysis) and for analysis of
CC gene expression. Proteins encoded by the nucleic acid sequences can be
CC used to screen for compounds that modulate gene expression, replication
CC and pathogenicity of *Listeria* (potential therapeutic agents), also for
CC treating infections by *Listeria*, and are useful as immunogens in
CC anti-*Listeria* vaccines.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic format
CC directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 3024 BP; 1114 A; 406 C; 590 G; 914 T; 0 other;

XX Query Match 3.9%; Score 41.6; DB 24; Length 3024;

XX Best Local Similarity 50.5%; Pred. No. 1.5;

XX Matches 101; Conservative 0; Mismatches 99; Indels 0; Gaps 0;

XX 43 TATCTTCCTTCTTGGAATGGAATGGCAAGATTAAATGATCAATTGATCAAGAT 102

XX 493 TATATATTTTCAAAATTTGAAATGAAATATACACACATTTCTTAATGCGTTGAGCAA 552

XX 103 GTTTGGCAACTCAGGTTCCGATATATATCGAATATACCGAATTTGATTGTTACAGT 162

XX 553 GTTGGCATATACAGATTTATGTCAGAGTCAAGATACAGATTTAGATATATTACG 612

XX 163 GATGCTAGTCTTCAAGTCTTCCAGATGTTTCTCAACTGATGATTTCTTCATAT 222

XX 613 GCTGCATATCAAGAAATGCCCCAGAGTTATCTATATTTGCGTTATATATTGTTT 672

XX 223 CTGTTCTCTTGGGATGAT 242

XX 673 GGTGAATTTTGTCTGATTT 692

XX Db

XX RESULT 6

XX ABL34073

XX ID ABL34073 standard; DNA; 40862 BP.

XX AC ABL34073;

XX DT 26-MAR-2002 (first entry)

XX DE Human immune system associated gene SEQ ID NO: 2046.

XX Human; immune system disease; cytosine methylation; antiasthmatic;

XX antiarteriosclerotic; antiamebic; cyostatic; nootropic;

XX neuroprotective; anti-HIV; anticonvulsant; ophthalmological;

XX antineumatic; antiarthritic; antidiabetic; antipsoriatic;

XX antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;

XX acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;

XX neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;

XX gene; ds.

XX OS *Homo sapiens*.

XX PN WO200200928-A2.

XX PD 03-JAN-2002.

XX PF 02-JUL-2001; 2001WO-EP07537.

XX PR 30-JUN-2000; 2000DE-1032529.

XX PR 01-SEP-2000; 2000DE-1043826.

XX PA (EPIC-) EPICENOMICS AG.

XX PI Olek A, Piepenbrock C, Berlin K;

XX WPI; 2002-130909/17.

XX DR

XX PT

XX PT

XX PT

XX PS

XX

XX

XX

XX

PT Nucleic acid comprising fragment of chemically modified gene, useful
PT for diagnosis and treatment of diseases associated with abnormal
PT cytosine methylation -
XX
PS Claim 1; SEQ ID NO 2046; 32pp + Sequence Listing; German.
XX
CC The present invention provides a number of human immune system associated
CC genes which are modified by the methylation of cytosines. The sequences
CC can be used in the diagnosis and treatment of immune system disorders,
CC including eye diseases such as retinopathy, neovascular glaucoma and
CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
CC leukemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
CC rheumatoid arthritis, psoriasis and inflammatory/intestinal bowel
CC diseases. The present sequence is a gene of the invention.
XX
SQ Sequence 40862 BP; 14301 A; 414 C; 7426 G; 18721 T; 0 other:
Query Match 3.9%; Score 41.6; DB 24; Length 40862;
Best Local Similarity 51.6%; Pred. No. 3.4;
Matches 95; Conservative 0; Mismatches 89; Indels 0; Gaps 0;
QY 118 GTTCCGATATATATGATATACCAATTTGATTTTCACTGATGCTGATCTTCAA 177
DB 18070 GGTTCGCAAAATTTTATTTTATTTGATGTTTGTATTTGATGATGATTTT 18129
QY 178 AGTCTTCAGATGTTTCTCACTGATGATGATTTCTTCATATCTTTGGGT 237
DB 18130 TGTGTGTAGAGATTTTATTTATTTATTTATTTATTTATTTGTTTGT 18189
QY 238 GGTGTCAAGATTTGATGATGATGATGATGATGATGATGATGATGATGAT 297
DB 18190 TATTTGTTTGTGTTTGTGATGATGATGATGATGATGATGATGATGAT 18249
QY 298 GTT 301
DB 18250 GTT 18253

RESULT 7
AB069244
ID AB069244 standard; DNA; 81905 BP.
XX
AC AB069244;
XX
DT 29-AUG-2002 (first entry)
XX
DE Listeria innocua DNA sequence #683.
XX
KW Antibacterial; Listeria; food contamination; mutational analysis;
XX infection; ds.
XX
OS Listeria innocua.
XX
PN WO200228891-A2.
XX
PD 11-APR-2002.
XX
PF 04-OCT-2001; 2001WO-FR03061.
XX
PR 04-OCT-2000; 2000FR-0012697.
XX
PA (INSP) INST PASTEUR.
XX (CNRS) CNRS CENT NAT RECH SCI.
XX
PI Kunst F, Glaeser P;
XX
DR WPI; 2002-332479/37.
XX
PT New genomic sequences from Listeria species, useful for detection,
PT treatment and prevention of infection, also related polypeptides,
PT antibodies and modulators -
XX
PS Claim 5; SEQ ID 2057; 180pp; French.

XX
CC The present invention relates to nucleic acid sequences
CC (AB067198-AB071212) from Listeria sp. The sequences are useful as probes
CC and primers for identification and/or detection of Listeria (e.g. as
CC contaminants in foods, or mutational analysis) and for analysis of
CC gene expression. Proteins encoded by the nucleic acid sequences can be
CC used to screen for compounds that modulate gene expression, replication
CC and pathogenicity of Listeria (potential therapeutic agents), also for
CC treating infections by Listeria, and are useful as immunogens in
CC anti-Listeria vaccines.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic format
CC directly from WIPO at http://wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 81905 BP; 26162 A; 14249 C; 14844 G; 26650 T; 0 other:
Query Match 3.9%; Score 41.6; DB 24; Length 81905;
Best Local Similarity 50.5%; Pred. No. 4.2;
Matches 101; Conservative 0; Mismatches 99; Indels 0; Gaps 0;
QY 43 TATCTCTCTTGGAATGAAATGGCAAGATTAAATTGATCAATTTGATCAAGAT 102
DB 24162 TATATATTTTCAAAATTTGAAATGAAATGAAATGAAATGAAATGAAATG 24221
QY 103 GTTTGGCAACTGAGTTTCCGATATATATGAAATATACCAATTTGATTTACAT 162
DB 24222 GTTTGGCATATATGATGATTTATGTCATGAGTCCAGAAATGATTTATTTACG 24281
QY 163 GATGCTAGTCTTCAAAAGCTTTTCCAGATGTTTCTCAACGATAGCATTTCTTCAAT 222
DB 24282 GCTGCATATCAAGAGAAATGCCAGAGATTATCTACTATATTTGTTATTAATTTGTTT 24341
QY 223 CTGTTCTTTGGGTGTT 242
DB 24342 GGTGAATTTTGTCTGATTT 24361

RESULT 8
AB067198
ID AB067198 standard; DNA; 82689 BP.
XX
AC AB067198;
XX
DT 29-AUG-2002 (first entry)
XX
DE Listeria innocua plasmid DNA sequence.
XX
KW Antibacterial; Listeria; food contamination; mutational analysis;
XX infection; ds.
XX
OS Listeria innocua.
XX
PN WO200228891-A2.
XX
PD 11-APR-2002.
XX
PF 04-OCT-2001; 2001WO-FR03061.
XX
PR 04-OCT-2000; 2000FR-0012697.
XX
PA (INSP) INST PASTEUR.
XX (CNRS) CNRS CENT NAT RECH SCI.
XX
PI Kunst F, Glaeser P;
XX
DR WPI; 2002-332479/37.
XX
PT New genomic sequences from Listeria species, useful for detection,
PT treatment and prevention of infection, also related polypeptides,
PT antibodies and modulators -
XX
PS Claim 5; SEQ ID 11; 180pp; French.

CC The present invention relates to nucleic acid sequences
CC (ABQ67188-ABQ71212) from *Listeria* sp. The sequences are useful as probes
CC and primers for identification and/or detection of *Listeria* (e.g. as
CC contaminants in foods, or mutational analysis) and for analysis of
CC gene expression. Proteins encoded by the nucleic acid sequences can be
CC used to screen for compounds that modulate gene expression, replication
CC and pathogenicity of *Listeria* (potential therapeutic agents), also for
CC treating infections by *Listeria*, and are useful as immunogens in
CC anti-*Listeria* vaccines.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic format
CC directly from WIPO at ftp.wipo.int/pub/published_ptc_sequences.
XX Sequence 82689 BP; 26448 A; 14348 C; 14977 G; 26916 T; 0 other;
SQ
Query Match 3.9%; Score 41.6; DB 24; Length 82689;
Best Local Similarity 50.5%; Pred. No. 4.2;
Matches 101; Conservative 0; Mismatches 99; Indels 0; Gaps 0;
0 43 TATCTCTCTTCTGGAAATGAAATGCGAAGTAAATGATCAATTTGATCAAGAT 102
10877 TATATATTTTCAAAATTTGAAATGAAATGACACACAACTTTAAATGCGTTGAGCAA 10936
Qy 103 GTTTTGGCAACTCAGGTTTCCGATATATATCGAATATACCAATTTGATTACAGT 162
10937 GTTTGGCAATATCAGATATATGATGAGGTACGAAATACAGATTAATTAATTAACG 10996
Qy 163 GATGCTAGTCTTCAAGCTTCCAGATGTTTCTCACTGATAGCATTTCTTCATAT 222
10997 GCTGCAATATCAAGAAATGCCCGAGCTTATCTACTATATGCGTTATATATTTGTTT 11056
Qy 223 CTGTGTTCTTGGGTGGTGT 242
Db 11057 GGTGAATTTTGTCTGATTT 11076
RESULT 9
ID AAA70156 standard; DNA; 7491 BP.
XX AAA70156;
AC
XX
DT 07-NOV-2000 (first entry)
XX
DE Plasmodium falciparum chromosome 2 related DNA sequence SEQ ID NO:289.
XX
KM Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;
KM antimalarial; malaria; protozoicide; infection; insecticide; ds.
XX
XX Plasmodium falciparum.
XX
PN WO200025728-A2.
XX
PD 11-MAY-2000.
XX
PE 05-NOV-1999; 99WO-US26796.
XX
PR 05-NOV-1998; 98US-0107131.
XX
PA (HOFF/) HOFFMAN S.
PA (CARU/) CARUCCI D.
PA (GARD/) GARDNER M.
PA (VENT/) VENTER J C.
XX
PI Hoffman S, Carucci D, Gardner M, Venter JC;
XX
DR WPI; 2000-365347/31.
XX
PT Proteins encoded by chromosome 2 of the human malarial parasite,
PT Plasmodium falciparum, useful as antimalarial vaccines and in the
PT diagnosis of P.falciparum infection -
XX
XX Disclosure; Page 496-497; 577pp; English.

XX
CC The present invention describes proteins and their fragments (I) encoded
CC by chromosome 2 of the human malarial parasite, Plasmodium falciparum.
CC Also described are: (I) nucleotide sequences (II) encoding (I); and (2)
CC vaccines against P. falciparum infection comprising (I) or (II).
CC (I) and (II) are useful for the development of vaccines against
CC P. falciparum infection. (I) and polyclonal antisera or a monoclonal
CC antibody raised to immunogens comprising the sequences of (I), are
CC useful in the detection of infection with P. falciparum. Furthermore,
CC (I) (especially when they are rifins or secreted or membrane proteins)
CC can aid the identification of drugs to treat or prevent P. falciparum
CC infection, or they can be used to identify drug resistance in
CC P. falciparum. Sequencing of the plasmodium chromosome 2 and the
CC subsequent identification of proteins encoded by it will help to expand
CC our understanding of parasite biology, a process hampered by the
CC complexity of the parasitic life cycle, and provide new targets for
CC vaccine and drug development. Parasite resistance to drugs and mosquito
CC resistance to insecticides have led to a resurgence of malaria in many
CC parts of the world, and there is a pressing need for vaccines and new
CC drugs. AAA70078 to AAA70287 and AAB18144 to AAB18352 represent nucleotide
CC and protein sequences given in the present invention, but which are not
CC specifically mentioned within the specification.
XX
SQ Sequence 7491 BP; 3252 A; 787 C; 964 G; 2488 T; 0 other;
Query Match 3.9%; Score 41.2; DB 21; Length 7491;
Best Local Similarity 46.0%; Pred. No. 2.5;
Matches 139; Conservative 0; Mismatches 163; Indels 0; Gaps 0;
Qy 580 ATCACTGAAATATTAACACATTTGGATTCAACAGTGTGCTGCAAGAAATAT 639
373 ATTTATATATTAATTAAGAAACTTTTGTGAACCTTTTGAATTTTGGGATAT 432
Db 640 ATTAACTCTGTAAGCAAGTAAACCAATTTCAACTACTATGTCGATATAGTAAAGT 699
433 ATATTTATATGAAATTAATATATTTTAAATATATACATTAATGTTGATTAACAAA 492
Qy 700 ATGTGAAATTAAGATATATCAGATCAACATATCAATGTTCACATTTGTTACC 759
493 AAGTAAATTAAGAAATTAAGAAATTAAGAAATTAAGAAATTAAGAAATTAAGAAAT 552
Qy 760 TCTGATTAAGCATATCGTTTATGATGATGTTTGGCTTTGAATTCATTTGTGATGAA 819
553 TATATATATATATTTTATGTTCAATATTTTATTTATATGATGATGAAAGAAAAA 612
Qy 820 AAGCAAAAGTGTGCTGTTGAAGTTATCAAGAAATTTATGACTTATGATTTTGGAC 879
613 AAAAGAAATGATATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 672
Qy 880 TT 881
Db 673 GT 674
RESULT 10
ID ABAK40092 standard; DNA; 9728 BP.
XX ABAK40092;
AC
XX
DT 21-MAY-2002 (first entry)
XX
DE Human chemically pretreated gene sequence #87 strand 2.
XX
KM Human; de; bisulphite treatment; CpG; DNA methylation; cancer; tumour;
KM cytochrome; ALDH6; CYP11A; CYP11B1; CYP3A3; DPYD; EPHX2; OCLN; TXNRD1;
KM UGT8; MRP; pharmacogenomics; SNP; single nucleotide polymorphism.
XX
XX Homo sapiens.
XX
PN WO200202806-A2.
XX
PD 10-JAN-2002.

XX 29-JUN-2001; 2001WO-EP07470.
 PF 30-JUN-2000; 2000DE-1032529.
 PR 01-SEP-2000; 2000DE-1043826.
 XX (EPIC-) EPIGENOMICS AG.
 PA
 PI Olek A, Piepenbrock C, Berlin K;
 DR WPI; 2002-154757/20.
 PT New nucleic acid, oligonucleotides and peptide nucleic acid-oligomers,
 with pharmacogenomics and for therapy of diseases e.g. cancer -
 PS Claim 1; SEQ ID No 174; 24pp; English.

CC The invention relates to a nucleic acid comprising a sequence at
 CC least 18 bases in length of a segment of the chemically pretreated DNA
 CC of genes associated with pharmacogenomics according to one of the
 CC sequences of the genes ALDH6 (NM_000693), CYP11A (NM_000781), CYP11B1
 CC (NM_000487), CYP3A3 (NM_000776 and NM_017460), DPYD (NM_000110), EPHX2
 CC (NM_001979), OCIN (NM_002538), TXNRD1 (NM_003330), UGT8 (NM_003360),
 CC MRP (NM_004996, NM_019900, NM_019901, NM_019902, NM_019862, NM_019898,
 CC NM_019899) and their complementary sequences, or a sequence (SI) chosen
 CC from 87 sequences and their complements. The chemical pretreatment
 CC is bisulphite treatment to convert cytosines (but not methyl-cytosines)
 CC into uraciles. Also included are an oligomer (II) in particular an
 CC oligonucleotide or a peptide nucleic acid (PNA)-oligomer, comprising in
 CC each case at least one base sequence having a length of 9 nucleotides
 CC which hybridises to or is identical to a chemically pretreated DNA of
 CC genes associated with pharmacogenomics and their complements, arranged in
 CC an array for analysing diseases associated with the methylation state
 CC (PPG) and/or detecting SNPs (single nucleotide polymorphisms)
 CC of the 87 sequences. The oligomers may also be used as PCR primers.
 CC The set of 87 nucleic acids and their complements is useful for diagnosis
 CC and therapy of solid tumours and cancer. The present sequence
 CC represents one the 87 DNA sequences or its complement.
 CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 9728 BP; 2972 A; 124 C; 1890 G; 4742 T; 0 other;

XX Query Match 3.8%; Score 41; DB 24; Length 9728;

XX Best Local Similarity 48.1%; Pred. No. 3;

XX Matches 116; Conservative 0; Mismatches 125; Indels 0; Gaps 0;

OY 635 AATATATTAAGCTGTGAAGCAAGTAACCAATTTCAACTACTAGTCGATATAGT 694
 Db 8677 ATTTTATTAAGCTTTTATATATTTGGAGAAAGAAAAATTTTATTTTGAATA 8736
 OY 695 AAGATATGTGAATTAAGATATCATCAGAGATCAACAATAGTTCACTCATTTG 754
 Db 8737 AAGTTTAATTAAGAAAAAAGATATATATATAGTATTTTAGATAGTATTTTAG 8796
 OY 755 GTACCTGTGATTAAGCCATACCTTATATCTAGTGTGGCTTGAATTCATTTGCTG 814
 Db 8797 GAAATGTAATTAATGTGAGATGATATATTTTGAAGTTTATATATATATATTTATT 8856
 OY 815 ATGAAAAGCAAAAGGTTGCTGTGAAGTATCAAGAATTTTATGACTTAATCTTAGTT 874
 Db 8857 GTAGTAGTAATTAAGATAGTAGCTTATTTATTAATATATATATATATATGATTTGATT 8916
 OY 875 T 875
 Db 8917 T 8917

RESULT 11
 ABL33903

ID ABL33903 standard; DNA; 9728 BP.

XX ABL33903;

XX 26-MAR-2002 (first entry)

XX Human immune system associated gene SEQ ID NO: 1876.

XX Human; immune system disease; cytosine methylation; antiasthmatic;
 KW antiarteriosclerotic; antianaemic; cytosatic; noctropic;
 KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;
 KW antirheumatic; antiarthritic; antidiabetic; antipsoriatic;
 KW antinflammatory; cancer; eye disease; arteriosclerosis; anaemia;
 KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;
 KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease;
 KW gene; ds.

XX Homo sapiens.

XX WO200200928-A2.

XX 03-JAN-2002.

XX 02-JUL-2001; 2001WO-EP07537.

XX 30-JUN-2000; 2000DE-1032529.

XX 01-SEP-2000; 2000DE-1043826.

XX (EPIC-) EPIGENOMICS AG.

XX Olek A, Piepenbrock C, Berlin K;

XX WPI; 2002-130909/17.

XX Nucleic acid comprising fragment of chemically modified gene, useful
 PT for diagnosis and treatment of diseases associated with abnormal
 PT cytosine methylation -

XX Claim 1; SEQ ID NO 1876; 32pp + Sequence Listing; German.

XX The present invention provides a number of human immune system associated
 CC genes which are modified by the methylation of cytosines. The sequences
 CC can be used in the diagnosis and treatment of immune system disorders,
 CC including eye diseases such as retinopathy, neovascular glaucoma and
 CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid
 CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,
 CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel
 CC diseases. The present sequence is a gene of the invention.

XX SQ Sequence 9728 BP; 2972 A; 124 C; 1890 G; 4742 T; 0 other;

XX Query Match 3.8%; Score 41; DB 24; Length 9728;

XX Best Local Similarity 48.1%; Pred. No. 3;

XX Matches 116; Conservative 0; Mismatches 125; Indels 0; Gaps 0;

OY 635 AATATATTAAGCTGTGAAGCAAGTAACCAATTTCAACTACTAGTCGATATAGT 694
 Db 8677 ATTTTATTAAGCTTTTATATATTTGGAGAAAGAAAAATTTTATTTTGAATA 8736
 OY 695 AAGATATGTGAATTAAGATATCATCAGAGATCAACAATAGTTCACTCATTTG 754
 Db 8737 AAGTTTAATTAAGAAAAAAGATATATATATAGTATTTTAGATAGTATTTTAG 8796
 OY 755 GTACCTGTGATTAAGCCATACCTTATATCTAGTGTGGCTTGAATTCATTTGCTG 814
 Db 8797 GAAATGTAATTAATGTGAGATGATATATTTTGAAGTTTATATATATATTTATT 8856
 OY 815 ATGAAAAGCAAAAGGTTGCTGTGAAGTATCAAGAATTTTATGACTTAATCTTAGTT 874
 Db 8857 GTAGTAGTAATTAAGATAGTAGCTTATTTATTAATATATATATATATGATTTGATT 8916
 OY 875 T 875

Db 2630 AAAACCCCTTAAATTTAAAAATTTAAATCTTTAAATTAACCAATTAACATTAACCA 2571
QY 479 TTTTCGATTTGTATCAACAATTTACTCCAAATCATCATCATGAGCTGATGATTAACAAG 538
Db 2570 TATTTTAAATTCATTTATTTATTTATTTAAATATTAACATATTTATTTATTTAAAAA 2511
QY 539 CCTGTGATCTTCCCAATCTGGTGACCAAGTCAACAGATATCACTCAAAAAATATAGAA 598
Db 2510 ATCCCCATCTTTTAAATTAATCATTTAAATATTTTACAAATTAATTAATTAATATAT 2455
QY 599 CCATTTGGATTCACAGCTTGTGCTCTCAAGAGATATATTAATCTCTTAAGCAAG 658
Db 2454 -----TTCAAAATTTATTTCCCAATTAATAAATAATTAATTAATAAATAATC 2405
QY 659 GTAACCAATTTCAAACTACTATGCGATATAGTGAAGATGTGGAATTAAGATA 718
Db 2404 GAAAAAATATTTATTAATTAACAAATAATCACTATTAATTTATTTTAAACCACTATTA 2345
QY 719 TCATCAGAGATCAACAATTAACATGTTTC 745
Db 2344 AAAACATTAATACTTCACTAATCTATTC 2318

RESULT 14
ABL34624/c
ID ABL34624 standard; DNA; 15518 BP.
XX
AC ABL34624;
XX
DT 26-MAR-2002 (first entry)
XX
DE Human metastasis associated gene SEQ ID NO: 177.
XX
KM Metastasis associated gene; cytostatic; gene therapy; cancer;
XX
KM cytosine methylation; gene; ds.
XX
OS Homo sapiens.
XX
PN M0200177376-A2.
XX
PD 18-OCT-2001.
XX
PF 06-APR-2001; 2001MO-EP03970.
XX
PR 06-APR-2000; 2000DE-1019058.
PR 07-APR-2000; 2000DE-1019173.
PR 30-JUN-2000; 2000DE-1032529.
PR 01-SEP-2000; 2000DE-1043826.
XX
XX (EPIC-) EPIGENOMICS AG.
XX
PI Olek A, Piepenbrock C, Berlin K;
XX
DR WPI; 2002-010922/01.
XX
PT New nucleic acid derived from chemically treated metastasis genes,
XX also for diagnosis of cancers by analysis of cytosine methylation,
XX PT also for treatment -
XX
PT Claim 1; SEQ ID NO 177; 23pp + Sequence Listing; English.
XX
PS
CC The present invention provides a number of human metastasis associated
CC genes which are modified by cytosine methylation. The sequences can be
CC used in the diagnosis and treatment of cancer. The present sequence is
CC one of the genes of the invention.
XX
SQ Sequence 15518 BP; 3047 A; 362 C; 4183 G; 7926 T; 0 other;

Query Match. 3.8%; Score 41; DB 24; Length 15518;
Best Local Similarity 48.3%; Pred. No. 3.5; Indels 14; Gaps 1;
Matches 18; Conservative 0; Mismatches 155; Indels 14; Gaps 1;
QY 419 AAAAGTTGTTATGACAAATGTTATCCAGATGTTGCTTCTTCTTCAAG 478

Db 2630 AAAACCCCTTAAATTTAAAAATTTAAATCTTTAAATTAACCAATTAACATTAACCA 2571
QY 479 TTTTCGATTTGTATCAACAATTTACTCCAAATCATCATCATGAGCTGATGATTAACAAG 538
Db 2570 TATTTTAAATTCATTTATTTATTTTAAATATTAACATATTTATTTATTTAAAAA 2511
QY 539 CCTGTGATCTTCCCAATCTGGTGACCAAGTCAACAGATATCACTCAAAAAATATAGAA 598
Db 2510 ATCCCCATCTTTTAAATTAATCATTTAAATATTTTACAAATTAATTAATTAATATAT 2455
QY 599 CCATTTGGATTCACAGCTTGTGCTCTCAAGAGATATATTAATCTCTTAAGCAAG 658
Db 2454 -----TTCAAAATTTATTTCCCAATTAATAAATAATTAATTAATAAATAATC 2405
QY 659 GTAACCAATTTCAAACTACTATGCGATATAGTGAAGATGTGGAATTAAGATA 718
Db 2404 GAAAAAATATTTATTAATTAACAAATAATCACTATTAATTTATTTTAAACCACTATTA 2345
QY 719 TCATCAGAGATCAACAATTAACATGTTTC 745
Db 2344 AAAACATTAATACTTCACTAATCTATTC 2318

RESULT 15
AAF58662/c
ID AAF58662 standard; DNA; 1537 BP.
XX
AC AAF58662;
XX
DT 27-APR-2001 (first entry)
XX
DE Porcine myostatin gene promoter.
XX
KM Porcine; pig; myostatin; promoter; cytostatic; anti-HIV; neuroprotective;
KM immunomodulator; cancer; muscular dystrophy; spinal cord injury;
KM neurodegenerative disorder; congestive obstructive pulmonary disease;
KM COPD; amyotrophic lateral sclerosis; acquired immunodeficiency syndrome;
KM AIDS; cachexia; ageing; ds.
XX
OS Sus scrofa.
XX
PN EP1072680-A1.
XX
PD 31-JUN-2001.
XX
PF 27-JUL-2000; 2000EP-0306396.
XX
PR 30-JUL-1999; 99US-0146540.
XX
XX (PFIZ) PFIZER PROD INC.
XX
PI Findly RC;
XX
DR WPI; 2001-170964/18.
XX
PT Novel myostatin gene promoter for regulating expression of heterologous
XX genes in cells or animals, for identifying compounds that inhibit its
XX activity or expression, which are useful for treating muscle disorders
XX -
XX
PS Disclosure; Page 12-13; 24pp; English.
XX
CC The present sequence was used for comparison studies with a myostatin
CC gene promoter sequence isolated from Mus musculus. Regulating the
CC transcription activities or expression of the isolated promoter is useful
CC for treating muscle associated disorders such as cancer, muscular
CC dystrophy, spinal cord injury, neurodegenerative disorders, traumatic
CC injury, congestive obstructive pulmonary disease (COPD), amyotrophic
CC lateral sclerosis, acquired immunodeficiency syndrome (AIDS), cachexia
CC or ageing. The promoter is useful for modulating the expression of the
CC myosin gene for producing animal food products having increased muscle
CC and protein content and reduced fat and cholesterol content, for

CC regulating expression of a heterologous gene in cells or animals, to
CC engineer host cells, to screen for compounds that inactivate or inhibit
CC its transcription and expression, for inhibiting its expression in cells
CC for the promotion of muscle growth, for diagnostic evaluation, genetic
CC testing and prognosis of a disease or disorder associated with myostatin
CC expression.

XX

SQ Sequence 1537 BP; 519 A; 259 C; 236 G; 523 T; 0 other;

SQ

Query Match 3.8%; Score 40.8; DB 22; Length 1537;

Best Local Similarity 49.1%; Pred. No. 1.9; Mismatches 0; Gaps 0;

Matches 108; Conservative 0; Indels 112; Gaps 0;

QY 799 AATTCCAATTGTCGATGAAAAGCAAAAGGTCGTTGAAGTATCAGAAATTATG 858

Db 1284 AATTAAATCTATGTAACATAATATTAACCTGACCTTAATATTTATTTAG 1225

QY 859 ACTAATCTTTAGTTGGACTGTTGGGCTCGGATTACTTCCAGCCACAGAT 918

Db 11224 TGAACCTTTAAATGTAATTAATTAAGTAATCTTTACACTTCATTAATAAGTT 1165

QY 919 GGTATGCTCATTTGGCTAAATCATCAACTTTATGCTCAATTGAGCCACAATTGAT 978

Db 1164 GAAATGCTTTTCTCTCAAAAAATTAATGATTTTACATTATTTAGACAAACATTGAG 1105

QY 979 GCCAAGGAAAGTAGAGTTTGGATGATGTTGACT 1018

Db 1104 GAAAAAAGAAATTATTTGCTGTTTAAATTAATTTTACT 1065

Search completed: January 27, 2003, 04:31:07
Job time : 467 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd

OM nucleic - nucleic search, using sw model

Run on: January 27, 2003, 04:18:00 ; Search time 52 Seconds
(without alignments)
6298.669 Million cell updates/sec

Title: US-09-675-509-3
Page: 1068

Sequence: 1 atgtccactccaaccaagac.....gtgtcttgagaccattcctt 1068

scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

number of hits satisfying chosen parameters: 8827224

```
Minimum DB seq length: 0
Maximum DB seq length: 20000000000
```

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

```
Database : Issued Patents NA:*
1 : /cgn2_6/prodata/1/ina/5A COMB. seg.*
2 : /cgn2_6/prodata/1/ina/5B COMB. seg.*
3 : /cgn2_6/prodata/1/ina/5A COMB. seg.*
4 : /cgn2_6/prodata/1/ina/5B COMB. seg.*
5 : /cgn2_6/prodata/1/ina/PCTUS COMB. seg.*
6 : /cgn2_6/prodata/1/ina/bckfil1e1. seg.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	54.2	5.1	7218	1	US-08-222-463-14	Sequence 14, Appl1
2	39.6	3.7	2287	1	US-08-222-619-1	Sequence 1, Appl1
3	39.6	3.7	2287	4	US-08-221-767-23	Sequence 23, Appl1
4	39.6	3.7	2287	5	PCT-US95-04075-1	Sequence 1, Appl1
5	38	3.6	1103	1	US-08-553-633-44	Sequence 4, Appl1
6	37.2	3.5	2334	4	US-09-071-035-361	Sequence 361, Appl1
7	36.6	3.4	19124	2	US-08-487-626B-13	Sequence 13, Appl1
8	35.8	3.4	825	4	US-09-134-001C-1569	Sequence 1569, Appl1
9	35.2	3.3	3706	4	US-08-913-159-9	Sequence 9, Appl1
10	35.2	3.3	16950	4	US-09-453-702B-165	Sequence 165, Appl1
11	35	3.3	1766	4	US-09-484-970B-132	Sequence 132, Appl1
12	35	3.3	2728	3	US-08-836-402B-7	Sequence 7, Appl1
13	35	3.3	5910	1	US-08-195-814-1	Sequence 1, Appl1
14	34.6	3.2	3471	1	US-07-876-280-29	Sequence 29, Appl1
15	34.6	3.2	3471	1	US-07-812-180A-1	Sequence 1, Appl1
16	34.6	3.2	3471	1	US-08-315-468-1	Sequence 1, Appl1
17	34.6	3.2	3471	1	US-07-941-650A-1	Sequence 1, Appl1
18	34.2	3.2	624	4	US-09-397-992A-3	Sequence 3, Appl1
19	34.2	3.2	624	4	US-09-397-992A-6	Sequence 6, Appl1
20	34	3.2	1750	4	US-09-276-531-14	Sequence 34, Appl1
21	33.8	3.2	1968	4	US-09-208-724-1	Sequence 1, Appl1
22	33.8	3.2	4182	1	US-07-973-257-1	Sequence 1, Appl1
23	33.8	3.2	11443	1	US-08-961-527-49	Sequence 49, Appl1
24	33.8	3.2	36159	4	US-09-749-588-3	Sequence 3, Appl1
25	33.6	3.1	978	4	US-09-232-938A-45	Sequence 45, Appl1
26	33.6	3.1	1071	4	US-08-887-534A-71	Sequence 71, Appl1
27	33.6	3.1	1756	2	US-08-702-598-1	Sequence 1, Appl1

C	28	3	3	4	3	1	985	4	US-08-642-306B-1	Sequence 1, App11
C	29	33	3	4	3	1	985	4	US-08-638-972B-1	Sequence 1, App11
C	30	33	4	3	3	1	985	4	US-08-771-21A-1	Sequence 1, App11
C	31	33	4	3	1	19950	4	US-08-661-527-35	Sequence 35, App11	
C	32	33	3	1	1	723	4	US-09-134-001C-1263	Sequence 1263, App11	
C	33	3	3	1	1	140	4	US-08-624-268-7	Sequence 7, App11	
C	34	3	3	1	1	140	5	PCT-US93-10442-7	Sequence 7, App11	
C	35	3	3	1	1	1610	5	US-08-689-108-7	Sequence 7, App11	
C	36	3	3	1	1	1610	5	PCT-US94-10358-7	Sequence 7, App11	
C	37	3	3	1	1	1664	4	US-07-863-169A-6	Sequence 6, App11	
C	38	3	3	1	1	1664	2	US-08-429-964-6	Sequence 6, App11	
C	39	3	3	1	1	1664	4	US-07-935-087-6	Sequence 6, App11	
C	40	3	3	1	1	1664	5	PCT-US93-08062-6	Sequence 6, App11	
C	41	3	3	1	1	2162	4	US-08-188-275A-1	Sequence 1, App11	
C	42	3	3	1	1	2162	4	US-09-351-198-1	Sequence 1, App11	
C	43	3	3	1	1	2566	4	US-09-113-426-1	Sequence 1, App11	
C	44	3	3	1	1	2566	4	US-08-624-268-19	Sequence 19, App11	
C	45	3	3	1	1	2566	5	PCT-US93-10442-19	Sequence 19, App11	

ALIGNMENTS

```

1      RESULT 1
2      US-08-232-463-14
3      ; Sequence 14, Application US/08232463
4      ; Patent No. 5670367
5      ;
6      ; GENERAL INFORMATION:
7      ;
8      ; APPLICANT: DORNER, F.
9      ; APPLICANT: SCHEIFLINGER, F.
10     ; APPLICANT: FALKNER, F. G.
11     ; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
12     ; NUMBER OF SEQUENCES: 52
13     ;
14     ; CORRESPONDENCE ADDRESS:
15     ; ADDRESSEE: Foley & Lardner
16     ; STREET: 1800 Diagonal Road, Suite 500
17     ; CITY: Alexandria
18     ; STATE: VA
19     ;
20     ; COUNTRY: USA
21     ; ZIP: 22313-0299
22     ;
23     ; COMPUTER READABLE FORM:
24     ; MEDIUM TYPE: Floppy disk
25     ; COMPUTER: IBM PC compatible
26     ; OPERATING SYSTEM: PC-DOS/MS-DOS
27     ; SOFTWARE: Patent in Release #1.0, Version #1.25
28     ;
29     ; CURRENT APPLICATION DATA:
30     ; APPLICATION NUMBER: US/08/232,463
31     ;
32     ; FILING DATE:
33     ; CLASSIFICATION: 435
34     ; PRIOR APPLICATION DATA:
35     ; APPLICATION NUMBER: US/07/935,313
36     ;
37     ; FILING DATE:
38     ; APPLICATION NUMBER: EP 91 114 300.6
39     ; FILING DATE: 26-AUG-1991
40     ;
41     ; ATTORNEY/AGENT INFORMATION:
42     ; NAME: BENT, Stephen A.
43     ; REGISTRATION NUMBER: 29,768
44     ; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
45     ; TELECOMMUNICATION INFORMATION:
46     ; TELEPHONE: (703)836-9300
47     ; TELEFAX: (703)683-4109
48     ;
49     ; TELEX: 899149
50     ; INFORMATION FOR SEQ ID NO: 14:
51     ;
52     ; SEQUENCE CHARACTERISTICS:
53     ;
54     ; LENGTH: 7218 base pairs
55     ; TYPE: nucleic acid
56     ; STRANDEDNESS: single
57     ; TOPOLOGY: linear
58     ;
59     ; IMMEDIATE SOURCE:
60     ; CLONE: pTZgpt-Fis
61     ;
62     ; US-08-232-463-14

```


TOPOLOGY: unknown
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 318..2117
FEATURE:
NAME/KEY: mat_peptide
LOCATION: 381..2114
FEATURE:
NAME/KEY: sig_peptide
LOCATION: 318..380
US-08-221-767-23

Query Match 3.7%; Score 39.6; DB 4; Length 2287;
Best Local Similarity 47.6%; Pred. No. 0.32;
Matches 117; Conservative 0; Mismatches 129; Indels 0; Gaps 0;

OY 237 TGGTGCAGAGTTGGATGATCATTTGGTTCGTGTTACTGGTGAATTCATAGTTT 296
D 116 TAGATATACATTTGTTGATGATGATTTAGTATGTTGCTCACAAGAGTCT 175
OY 297 TGTTCCTCAGGCCCTGTCATGTTCCGTTATGTTCCCAATACCTGTGCTC 356
DB 176 AAATCCATAGCTTTATATTCAGGCTACTTTATTTTGAAGCTCATTTCTATCA 235
OY 357 AAATTTTATTTGCTCTCACAAGATGTTACTCAACAGCATCTCCCTTAGAATGGC 416
DB 236 CTTTTCTATTTTACTCCATATTTGAGGCTCATTAATCCATTTTATTTCTTT 295
OY 417 TCAAAAGTTGTTATGACAAATTTGTTATCCAGATGTTGCTCTTCTAGTTCTTAC 476
DB 296 TGTAAATGTTGTTCTTCAAAAGATGAAGTAACTAACTTAACAGTTTATTTTCTT 355
OY 477 AGTTT 482
DB 356 GTTTT 361

RESULT 4

PCT-US95-04075-1
Sequence 1, Application PC/TUS9504075
GENERAL INFORMATION:
APPLICANT: AMGEN INC.
TITLE OF INVENTION: Afamin: A Human Serum Albumin-Like
TITLE OF INVENTION: Protein
NUMBER OF SEQUENCES: 33
CORRESPONDENCE ADDRESS:
ADDRESSEE: Amgen Center, Patent Operations/RRC
STREET: 1840 Dehavilland Drive
CITY: Thousand Oaks
STATE: California
COUNTRY: U.S.
ZIP: 91320-1789
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/04075
FILING DATE:
CLASSIFICATION:
INFORMATION FOR SEQ. ID NO. 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2287 base pairs
TYPE: nucleic acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 318..2117
FEATURE:

NAME/KEY: mat_peptide
LOCATION: 381..2114
FEATURE:
NAME/KEY: sig_peptide
LOCATION: 318..380
PCT-US95-04075-1

Query Match 3.7%; Score 39.6; DB 5; Length 2287;
Best Local Similarity 47.6%; Pred. No. 0.32;
Matches 117; Conservative 0; Mismatches 129; Indels 0; Gaps 0;

OY 237 TGGTGCAGAGTTGGATGATCATTTGGTTCGTGTTACTGGTGAATTCATAGTTT 296
D 116 TAGATATACATTTGTTGATGATGATTTAGTATGTTGCTCACAAGAGTCT 175
OY 297 TGTTCCTCAGGCCCTGTCATGTTCCGTTATGTTCCCAATACCTGTGCTC 356
DB 176 AAATCCATAGCTTTATATTCAGGCTACTTTATTTTGAAGCTCATTTCTATCA 235
OY 357 AAATTTTATTTGCTCTCACAAGATGTTACTCAACAGCATCTCCCTTAGAATGGC 416
DB 236 CTTTTCTATTTTACTCCATATTTGAGGCTCATTAATCCATTTTATTTCTTT 295
OY 417 TCAAAAGTTGTTATGACAAATTTGTTATCCAGATGTTGCTCTTCTAGTTCTTAC 476
DB 296 TGTAAATGTTGTTCTTCAAAAGATGAAGTAACTAACTTAACAGTTTATTTTCTT 355
OY 477 AGTTT 482
DB 356 GTTTT 361

RESULT 5

US-08-553-633A-4
Sequence 4, Application US/08553633A
GENERAL INFORMATION:
APPLICANT: SALMOND, GEORGE PC
APPLICANT: HOLDEN, MATTHEW TG
APPLICANT: COX, ANTHONY RJ
APPLICANT: THOMSON, NICHOLAS R
APPLICANT: MCGOWAN, SIMON J
TITLE OF INVENTION: PROCESS FOR ACTIVATING GENE EXPRESSION
TITLE OF INVENTION: IN BACTERIA
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: BIRCH, STEWART, KOLASCH AND BIRCH
STREET: PO BOX 747
CITY: FALLS CHURCH
STATE: VA
COUNTRY: USA
ZIP: 22040-0747
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/553,633A
FILING DATE: 01-APR-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: SVENSSON, LEONARD R
REGISTRATION NUMBER: 30,330
REFERENCE/DOCKET NUMBER: 1009-104
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)-205-8000
TELEFAX: (703)-205-8050
INFORMATION FOR SEQ. ID NO. 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1103 base pairs
TYPE: nucleic acid
STRANDEDNESS: single

TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: Serratia
US-08-553-633A-4

Query Match 3.6%; Score 38; DB 1; Length 1103;
Best Local Similarity 43.6%; Pred. No. 0.67;
Matches 170; Conservative 0; Mismatches 220; Indels 0; Gaps 0;

598 ACCATTGATTCACAGTGTGGCTCCCAAGAGATATATTAACCTCTGTAACCAA 657
DB 330 ACCATATGATATTAATTAATCTTTATTAATGATCTTTTACTTTTAAAGA 389
QY 658 GGTAAACCAATTCACATCTATGTCGATATAGTAAAGTATGTAATTAAGAT 717
DB 390 GTAACTCATTTTCTGCGATGATAGATCTGCTGATTAAGGTCGAAATGAAGAT 449
QY 718 ATATCAGAGATCAACATATCAATGTTCAACTGTTGTAACCTCTGTAAGCATGCTT 777
DB 450 GTTGCCATGTATCAAGGAGATGATCATCATGTAGTTACATTTGTTCTTCAGAC 509
QY 778 TATACATATGTTTGGTTGATTCATTTGTGTGATGAAAGCAAAAGTTCCTGT 837
DB 510 CATGATTAACATCTGGCATCTGACTATGCTAATGATGAAAAAATGATTTTGA 569
QY 838 GAAGTATCAAGATTTATGACTAATTAATTAATTTGACTTGGCTGCGATTA 897
DB 570 GATTTTATTAAGAACAGAGAAATGATTAACAATGTTTGTAGACTACTACAGAAAA 629
QY 898 ACTCTCCAGCAACAGATGTTATGCTCATTTGGCTAAATCATCAACTTTATGCT 957
DB 630 GCATGAAACATTAACCTTCGTTAAAGGTAAACGGCGCCTTGATGCTTCAAGT 689
QY 958 CAATTGAGCAACATTCGATGCAAGAA 987
DB 690 GCATGATTAACACCGTGAACAGAAAGTA 719

RESULT 6

US-09-071-035-361/C
Sequence 361, Application US/09071035
Patent No. 6448043
GENERAL INFORMATION:
APPLICANT: Gil H. Choi
TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides
NUMBER OF SEQUENCES: 496
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4MB storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/071.035
CLASSIFICATION:
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: A. Anders Brookes
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB369P2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512

INFORMATION FOR SEQ ID NO: 361:
SEQUENCE CHARACTERISTICS:
LENGTH: 2334 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
US-09-071-035-361

Query Match 3.5%; Score 37.2; DB 4; Length 2334;
Best Local Similarity 55.4%; Pred. No. 1.4;
Matches 72; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

QY 189 TGTTCCTCACTGATGATGATTTCTTCATATCTGTTCTTGGGCGGTCAAG 248
DB 243 TGTTCCTCACTGATGATGATTTCTTCATATCTGTTCTTGGGCGGTCTGC 184
QY 249 TTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 308
DB 183 TGTTCCTCACTGATGATGATTTCTTCATATCTGTTCTTGGGCGGTCAAG 124
QY 309 TGCCTCTGTC 318
DB 123 AACAGTTGTC 114

RESULT 7

US-08-487-826B-13
Sequence 13, Application US/08487826B
Patent No. 5993827
GENERAL INFORMATION:
APPLICANT: Sim, Kim L.
APPLICANT: Chitnis, Chetan
APPLICANT: Miller, Louis H.
APPLICANT: Peterson, David S.
APPLICANT: Su, Xin-zhan
APPLICANT: Wellens, Thomas E.
TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX
NUMBER OF SEQUENCES: 45
CORRESPONDENCE ADDRESS:
ADDRESSEE: Knodde Martens Olson & Bear
STREET: 620 Newport Center Drive 16th floor
CITY: Newport Beach
STATE: California
COUNTRY: US
ZIP: 92660
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/487,826B
FILING DATE: 10-SEP-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Israelien, Ned
REGISTRATION NUMBER: 29,655
REFERENCE/DOCKET NUMBER: NIH121.001CP1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 19124 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-487-826B-13

Query Match 3.4%; Score 36.6; DB 2; Length 19124;
 Best Local Similarity 44.8%; Pred. No. 3.9;
 Matches 141; Conservative 0; Mismatches 174; Indels 0; Gaps 0;

QY 203 ATAGCATTTCCCTCCATCTCTGTTTGGTGTGTCAGAGTTTGATGATCAT 262
 DB 17579 ATATCCATATTCAGATATATGAAATATCTAAATTAATTAATCTTTTGTGTGTCAT 17638

QY 263 TGGTTCGTGTTTACGTGATTTGATTTGTTTCCCTCAAGTCCCTGTCAATG 322
 DB 17639 ATATATTTTATATTTATTTGATATGTTTATTTATTTATTTATTTATTTATTTAT 17698

QY 323 GTTCCGTTATGTTTCCCAATATCTGTCTCAAACTTTTATGTCCTCACCAATG 382
 DB 17699 TTTTGGATATTTGTAATAAATATGTTTGTATATCATATTTATTTATTTATTTATA 17758

QY 383 GTACTCAACAGACATCTCCCTTTAGAAATGGCTCAAAAGTTGGTTAAGCAAAATG 442
 DB 17759 ATTTGCAACATGATTTTTTTTTTTCTTTCTTATTTGTATTTTTCATATTAATTTATA 17818

QY 443 TTATCCAGATGTGCTCTTCTTCAAGTTTTCAGAGTTTTCGATTTGATCAACATTTAC 502
 DB 17819 TATATATATGATTTATTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTAT 17878

QY 503 TCCATCATCATCAT 517
 DB 17879 GTATATTATGTTAT 17893

RESULT 8
 US-09-134-001C-1569
 ; Sequence 1569, Application US/09134001C
 ; Patent No. 6380370
 ; GENERAL INFORMATION:
 ; APPLICANT: Lynn Doucette-Stamm et al
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
 ; FILE REFERENCE: GTC-007
 ; CURRENT APPLICATION NUMBER: US/09/134,001C
 ; CURRENT FILING DATE: 1998-08-13
 ; PRIOR APPLICATION NUMBER: US 60/064,964
 ; PRIOR FILING DATE: 1997-11-08
 ; PRIOR APPLICATION NUMBER: US 60/055,779
 ; PRIOR FILING DATE: 1997-08-14
 ; NUMBER OF SEQ ID NOS: 5674
 ; SEQ ID NO 1569
 ; LENGTH: 825
 ; TYPE: DNA
 ; ORGANISM: Staphylococcus epidermidis
 US-09-134-001C-1569

Query Match 3.4%; Score 35.8; DB 4; Length 825;
 Best Local Similarity 45.5%; Pred. No. 2.3;
 Matches 127; Conservative 0; Mismatches 152; Indels 0; Gaps 0;

QY 766 AAGCATACCTTATATGATGTTTGGCTTGAATTCGAATTTGTGTGATGAAGCAAA 825
 DB 37 AAGGCATACAGCAAAAGATTTCTATGTTACAGCTTATGATTTTCTAGTCTTAAGCA 96

QY 826 AAGGTGCTGTGAAGTATCAAGATTTATGCTAATACTTTAGTTTGGACTGTG 885
 DB 97 GCACAAACAGCTGAATTTGATCATGATTTTGTAGAGATTTCTTAAGAAAGACAGTGT 156

QY 886 GGTCTGGATTAATCTCCAGCCACACAGATGTTGTCTATTTGGCTAATCATCA 945
 DB 157 GGATATATATGATCTGTTCAAGTTAATGAAACATATGATTTCAATGTTAGAGCTGTT 216

QY 946 AACTTTATGCTCAATGAGCAACAATTCGATCCCAAGAAAGTGAAGTTAGATTTTG 1005
 DB 217 AAAAGAGTGTCTTCAATACATTTATATGTTGTTATATGCTTAATAGGAGACTGTGTTTA 276

QY 1006 AGATGTGTGACTTGTCTAACAGAAAGTTAAGATTTGT 1044

DB 277 AGTATGAGAGAGATCTAAAAATGCACTTAAGCTTTAT 315

RESULT 9
 US-08-913-159-9/C
 ; Sequence 9, Application US/08913159
 ; Patent No. 6300109
 ; GENERAL INFORMATION:
 ; APPLICANT:
 ; TITLE OF INVENTION: Plasmid-derived type II
 ; TITLE OF INVENTION: restriction-modification systems from Lactococcus lactis
 ; NUMBER OF SEQUENCES: 14
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/913,159
 ; FILING DATE:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: DK 0179/95
 ; FILING DATE: 17-FEB-1995
 ; INFORMATION FOR SEQ ID NO: 9:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 3706 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: double
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: DNA (genomic)
 ; ANTI-SENSE: NO
 ; ORIGINAL SOURCE:
 ; ORGANISM: Lactococcus lactis subsp. cremoris
 ; STRAIN: W56
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: complement (422..2161)
 ; IDENTIFICATION METHOD: experimental
 ; OTHER INFORMATION: /codon_start= 422
 ; OTHER INFORMATION: /product= "LlABI methylase"
 ; OTHER INFORMATION: /evidence= EXPERIMENTAL
 ; OTHER INFORMATION: /gene= "ORF"
 ; OTHER INFORMATION: /number= 1
 ; OTHER INFORMATION: /standard_name= "Gene coding for LlABI methylase"
 ; OTHER INFORMATION: /label= m_1l1abi
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: 2464..3360
 ; IDENTIFICATION METHOD: experimental
 ; OTHER INFORMATION: /codon_start= 2464
 ; OTHER INFORMATION: /product= "LlABI endonuclease"
 ; OTHER INFORMATION: /evidence= EXPERIMENTAL
 ; OTHER INFORMATION: /gene= "ORF"
 ; OTHER INFORMATION: /number= 2
 ; OTHER INFORMATION: /standard_name= "Gene coding for LlABI
 ; OTHER INFORMATION: endonuclease"
 ; OTHER INFORMATION: /label= f_1l1abi
 US-08-913-159-9

Query Match 3.3%; Score 35.2; DB 4; Length 3706;
 Best Local Similarity 50.0%; Pred. No. 5.3;
 Matches 88; Conservative 0; Mismatches 88; Indels 0; Gaps 0;

QY 610 TCAACAGTTGTGCTCTCAAGAGATATATTACTGTGAAGCAAGTAACCAATT 669
 DB 1201 TCTACTGGTATGCTACTCAAAAGATTAACCTTATTTATAGACTATCTAATAAGAA 1142

QY 670 TCAACTACTATGCGATATGATGTAAGATGCTGGAATTAAGATATCATCAGAGAT 729
 DB 1141 GTTAATATGATTAAGAAATATTTGTGAAAGATTAAGACACACCTTTCTCATAGAG 1082

QY 730 CAACAATACATGTCAACTCATTTGTTGATGATGATGATGATGATGATGATGATGATGAT 785

Db 1081 AAGAAATACTATTCTATTATTAAGACTGAGAGAAAAAGTTAATCTAA 1026

RESULT 10

US-09-453-702B-166/C
; Sequence 166, Application US/09453702B
; Patent No. 6365723

GENERAL INFORMATION:

APPLICANT: Blatner, Frederick R.
Burland, Nicole T.
Perna, Nicole T.
Plunkett, Guy
Welch, Rod

TITLE OF INVENTION: No. 6365723el Sequences of E. coli O157
NUMBER OF SEQUENCES: 265
CORRESPONDENCE ADDRESS:

ADDRESSEE: Quarles & Brady
STREET: 1 South Pinckney Street
CITY: Madison
STATE: WI
COUNTRY: US
ZIP: 53701-2113

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.44MB storage
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Word Perfect 8.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/453,702B
FILING DATE: 03-Dec-1999
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 60/110,955
FILING DATE: 04-DEC-1998
ATTORNEY/AGENT INFORMATION:
NAME: Seay, Nicholas J.

REGISTRATION NUMBER: 27386
REFERENCE/DOCKET NUMBER: 960296,95017
TELECOMMUNICATION INFORMATION:
TELEPHONE: (608) 251-5000
TELEFAX: (608) 251-9166

INFORMATION FOR SEQ ID NO: 166:
SEQUENCE CHARACTERISTICS:
LENGTH: 16950
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)
SEQUENCE DESCRIPTION: SEQ ID NO: 166:
US-09-453-702B-166

Query Match 3.3%; Score 35.2; DB 4; Length 16950;
Best Local Similarity 54.7%; Pred. No. 8.7;

Matches 70; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

QY 342 ACAATACCTGCTCAACTTTTATTTGCTCTCACCAATGTAAGCATCTTC 401
Db 163 AGATTATATGCTTCATTAATATCTCAACGAGTAAGCTTGAATATACAC 104
QY 402 CCTTTAGAAATGGCTCAAAAGTTGGTTATGAACAAATGTTATCCAGATGGCTC 461
Db 103 CTATTAACAATCCGGTGTGAGATCTATTACCAATGAATGATTCCTCGCAGATATATC 44
QY 462 TTCTAGTT 469
Db 43 TTTCAGTT 36

RESULT 11

US-09-484-970B-132/C
; Sequence 132, Application US/09484970B
; Patent No. 6426186

GENERAL INFORMATION:

APPLICANT: Jones, Karen A.
APPLICANT: Volkmuth, Wayne
APPLICANT: Walker, Michael G.
TITLE OF INVENTION: BONE REMODELING GENES
FILE REFERENCE: PB-0014 US
CURRENT APPLICATION NUMBER: US/09/484,970B
CURRENT FILING DATE: 2000-01-18
NUMBER OF SEQ ID NOS: 172
SOFTWARE: PERL Program
SEQ ID NO 132

LENGTH: 1766
TYPE: DNA

ORGANISM: Homo sapiens
FEATURE:

NAME/KEY: misc feature
OTHER INFORMATION: Incyte ID No. 6426186 245065.1CB1
US-09-484-970B-132

Query Match 3.3%; Score 35; DB 4; Length 1766;
Best Local Similarity 57.9%; Pred. No. 4.7;
Matches 62; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

QY 609 TTCAACAGTGTGCTCTCAAAAGAGATATATTAACTCTGTAAGCAAGTAAACCAAT 668
Db 334 TTCTAAAGTCTCTTCAAGCTTCAAAAGCATGATTAATCTGTAGGCACTGAAAAAAC 275

QY 669 TTTAACTACTATGCGATATAGTAAGATATGTGAAATTAAGG 715
Db 274 TGAGATTATGAAGATGATCTCTTCAATATATATATATAAAGS 228

RESULT 12

US-08-836-402B-7/C
; Sequence 7, Application US/08836402B
; Patent No. 6063988

GENERAL INFORMATION:
APPLICANT: Rudiger Hain, Regina Fischer
TITLE OF INVENTION: DNA SEQUENCE AND ITS USE
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:

ADDRESSEE: SPRUNG KRAMER SCHAEFER & BRISCOE
STREET: 660 White Plains Road
CITY: Tarrytown
STATE: New York
COUNTRY: U.S.A.

ZIP: 10591-5144
COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 2.0 MB storage
COMPUTER: Gateway 2000
OPERATING SYSTEM: DOS
SOFTWARE: Microsoft Windows 98
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/836,402B
FILING DATE: 02-SEP-1997
CLASSIFICATION: 800

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 44 40 200.7 (Germany)
FILING DATE: 10-NOV-1994
ATTORNEY/AGENT INFORMATION:
NAME: Kurt G. Briscoe

REGISTRATION NUMBER: 33,141
REFERENCE/DOCKET NUMBER: Bayer 9840-KGB
TELECOMMUNICATION INFORMATION:
TELEPHONE: (914) 332-1700
TELEFAX: (914) 332-1844

TELEX:
INFORMATION FOR SEQ ID NO: 7:

SEQUENCE CHARACTERISTICS:
LENGTH: 2728 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear

INDIVIDUAL ISOLATE: LARVAE
DEVELOPMENTAL STAGE: LARVAE
HAPLOTYPE: LARVAE

NAME: Saliwanchik, David R.
REGISTRATION NUMBER: 31,794
REFERENCE/DOCKET NUMBER: M/S 104

```

; TELEFAX: 904-372-5800
; INFORMATION FOR SEQ ID NO: 29:
; SEQUENCE CHARACTERISTICS:

```

LENGTH: 3471 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Bacillus thuringiensis
STRAIN: kumamotoensis
INDIVIDUAL ISOLATE: PS50C
IMMEDIATE SOURCE:
CLONE: E. coli NM522(pMYC320) NRRL B-18769
US-07-876-280-29

Query Match 3.2% Score 34.6; DB 1; Length 3471;
Best Local Similarity 55.4%; Pred. No. 7.4;
Matches 67; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

Q 416 CTCAAAAGTTGTTATGACAAATGTTTATCCAGATGTCCTCTTCTAGTTCTTTCA 475
D 491 CTAGAAAGTTGTTCTTAATCAATTTATAGCTTAGATCTTAAGTTGTTCAATTC 550
Q 476 CAGTTTCGATTGTATCAACAATTACTCCATCATCTCATCGCTGCAGTTGATATCA 535
D 551 CATCTTTGCAGTATCCGACACGAGTACTATTATTAGCAGTATATGACAGGCTGTGA 610

Q 536 A 536
D 611 A 611

RESULT 15

US-07-812-180A-1

; Sequence 1, Application US/07812180A
; Patent No. 5366892

; GENERAL INFORMATION:

; APPLICANT: Foncerrada, Luis R

; APPLICANT: Payne, Jewel M

; APPLICANT: Sick, August J

; TITLE OF INVENTION: No. 5366892el Coleopteran-Active Bacillus

; TITLE OF INVENTION: thuringiensis isolate and a No. 5366892el Gene Encoding a

; NUMBER OF SEQUENCES: 2

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Roman Saliwanchik

; STREET: 2421 N.W. 41st Street, Ste A-1

; CITY: Gainesville

; STATE: FL

; COUNTRY: USA

; ZIP: 32606

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS 5.0

; SOFTWARE: Patent Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/07/812,180A

; FILING DATE: 19920102

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Saliwanchik, Roman

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 904-375-8100

; TELEFAX: 904-372-5800

; INFORMATION FOR SEQ ID NO: 1:

; LENGTH: 3471 base pairs

; TYPE: NUCLEIC ACID

; STRANDEDNESS: double

; TOPOLOGY: linear

; MOLECULE TYPE: DNA (genomic)

; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Bacillus thuringiensis
; STRAIN: kumamotoensis
; INDIVIDUAL ISOLATE: PS50C
; IMMEDIATE SOURCE:
; LIBRARY: LAMBDAGEM (TM) - 11 LIBRARY OF LUIS
; LIBRARY: FONCERRADA
; CLONE: 50C
US-07-812-180A-1

Query Match 3.2% Score 34.6; DB 1; Length 3471;
Best Local Similarity 55.4%; Pred. No. 7.4;
Matches 67; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

Q 416 CTCAAAAGTTGTTATGACAAATGTTTATCCAGATGTCCTCTTCTAGTTCTTTCA 475
D 491 CTAGAAAGTTGTTCTTAATCAATTTATAGCTTAGATCTTAAGTTGTTCAATTC 550
Q 476 CAGTTTCGATTGTATCAACAATTACTCCATCATCTCATCGCTGCAGTTGATATCA 535
D 551 CATCTTTGCAGTATCCGACACGAGTACTATTATTAGCAGTATATGACAGGCTGTGA 610

Q 536 A 536
D 611 A 611

Search completed: January 27, 2003, 05:36:48
Job time: 114 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Comugen Ltd.

OM nucleic - nucleic search, using SW model

Run on: January 27, 2003, 05:30:37 ; Search time 58 Seconds
(without alignments)
8272.669 Million cell updates/sec

Title: US-09-675-509-3

Perfect score: 1068
Sequence: 1 atgtccaccacacaaagac.....gtgtcttgagaccattcctt 1068

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Sequences: 396772 seqs, 224632407 residues

Number of hits satisfying chosen parameters: 793544

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database:

Published Applications NA:
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PC7_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PC7US_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	42.4	4.0	491	US-09-878-574-4301	Sequence 4301, App
2	40	3.7	392	US-09-960-352-11703	Sequence 11703, A
3	39.6	3.7	2280	US-10-044-090-450	Sequence 450, App
4	39.6	3.7	2282	US-09-880-107-2192	Sequence 2192, App
5	38.2	3.6	490	US-09-783-590-6019	Sequence 6019, Ap
6	37.6	3.5	536	US-09-834-975-317	Sequence 317, App
7	37.2	3.5	1517	US-09-070-927A-180	Sequence 180, App
8	36.6	3.4	1503841	US-09-946-807-1	Sequence 1, Appli
9	36.6	3.4	1503841	US-09-795-666-1	Sequence 1, Appli
10	36.6	3.4	1503841	US-09-795-666-1	Sequence 1, Appli
11	35.8	3.4	2000	US-09-938-842A-4858	Sequence 4858, Ap
12	35.8	3.4	116592	US-09-818-512-3	Sequence 512, Appli
13	35.6	3.3	2295	US-09-815-242-3894	Sequence 3894, Ap
14	35.6	3.3	2295	US-09-815-242-6496	Sequence 6496, Ap
15	35.2	3.3	371	US-09-796-692-2970	Sequence 2970, Ap
16	35.2	3.3	371	US-09-796-692-5606	Sequence 5606, Ap
17	35.2	3.3	543	US-09-764-847-471	Sequence 471, App
18	35.2	3.3	2550	US-09-764-864-318	Sequence 318, App
19	35.2	3.3	3631	US-09-764-864-738	Sequence 738, App

20	35.2	3.3	7569	US-09-764-855-333	Sequence 333, App
21	35.2	3.3	7569	US-09-764-855-334	Sequence 334, App
22	35	3.3	1908	US-09-815-242-4783	Sequence 4783, Ap
23	35	3.3	1911	US-09-815-242-8910	Sequence 8910, Ap
24	34.6	3.3	52354	US-09-742-311-3	Sequence 3, Appli
25	34.6	3.2	2108	US-09-070-927A-568	Sequence 568, App
26	34.2	3.2	280	US-09-815-242-1659	Sequence 1659, Ap
27	34.2	3.2	385	US-09-960-352-8405	Sequence 8405, Ap
28	34.2	3.2	624	US-09-971-843-3	Sequence 3, Appli
29	34.2	3.2	624	US-09-971-843-6	Sequence 6, Appli
30	34.2	3.2	846	US-09-815-242-8818	Sequence 8818, Ap
31	34.2	3.2	4780	US-09-962-436-287	Sequence 287, App
32	34	3.2	2000	US-09-938-842A-479	Sequence 479, Ap
33	34	3.2	8993	US-09-764-869-1747	Sequence 1747, Ap
34	34	3.2	18946	US-10-147-026-15	Sequence 15, Appli
35	33.8	3.2	281	US-09-294-093B-1962	Sequence 1962, Ap
36	33.8	3.2	1245	US-09-815-242-8762	Sequence 8762, Ap
37	33.8	3.2	8062	US-10-108-605-294	Sequence 294, App
38	33.8	3.2	36159	US-10-135-687-3	Sequence 3, Appli
39	33.8	3.2	335913	US-09-754-853A-2	Sequence 2, Appli
40	33.8	3.2	335913	US-09-754-853A-3	Sequence 3, Appli
41	33.6	3.1	285	US-09-878-574-13954	Sequence 13954, A
42	33.6	3.1	23041	US-09-764-864-1753	Sequence 1753, Ap
43	33.6	3.1	513509	US-09-754-853A-4	Sequence 4, Appli
44	33.4	3.1	435	US-09-974-300-8049	Sequence 8049, Ap
45	33.4	3.1	435	US-09-974-300-8056	Sequence 8056, Ap

ALIGNMENTS

RESULT 1
US-09-878-574-4301
Sequence 4301, Application US/09878574
Patent No. US20020110548A1
GENERAL INFORMATION:
APPLICANT: Byrum, Joseph R.
APPLICANT: La Rosa, Thomas J.
APPLICANT: Thompson, Michael D.
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
FILE REFERENCE: 38-21 (15401) B
CURRENT APPLICATION NUMBER: US/09/878,574
PRIORITY FILING DATE: 2001-12-21
PRIORITY APPLICATION NUMBER: 09/333,535
NUMBER OF SEQ ID NOS: 15775
SEQ ID NO 4301
LENGTH: 491
TYPE: DNA
ORGANISM: Glycine max
FEATURE:
NAME/KEY: unsure
LOCATION: (1)..(491)
OTHER INFORMATION: unsure at all n locations
OTHER INFORMATION: Clone ID: LIB3028-014-Q1-B1-A11
US-09-878-574-4301
Query Match
Best local Similarity 47.8%; Pred. No. 0.24;
Matches 121; Conservative 0; Mismatches 132; Indels 0; Gaps 0;
QY 118 GTTTCGAGATATATATTCGATATACCGATTTGATTTACGATGATGCTTCA 177
DB 57 GTTGTAACTGATGTTTCGATTTGATTTGATTTGATTTGATTTGATTTGATTT 116
QY 178 AGCTTCGAGATTTTTCGACGATGATGATTTTCTTCATCTGTTGGGT 237
DB 117 TTTCTGTTGTTTGTGTTGTTGTTGTTTGTGTTTGTGTTTGTGTTTGTGTTT 176
QY 238 GGTGTCAAGATTTGATGATGATGATGATGATGATGATGATGATGATGATGAT 297
DB 177 GGTGTTGTTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTT 236

Db 296 TGTAAAGTGTCTCAAGAGTAACTAAACCTTACAGTTTATTTTCTT 355
 Oy 477 AGTTT 482
 Db 356 GTTTT 361

RESULT 5

US-09-783-590-6019
 / Sequence 6019 Application US/09783590
 / Patent No. US20020110850A1
 / GENERAL INFORMATION:
 / APPLICANT: Dillon, Patrick J.
 / APPLICANT: Haseltine, William A.
 / APPLICANT: Li, Haodong
 / APPLICANT: Rosen, Craig A.
 / APPLICANT: Ruden, Steven M.
 / TITLE OF INVENTION: Human Genes, Sequences, and Expression Products 16.2
 / PARENT REFERENCE: PO-16.2C1
 / CURRENT APPLICATION NUMBER: US/09/783,590
 / CURRENT FILING DATE: 2000-02-15
 / PRIOR APPLICATION NUMBER: 08/420,856
 / PRIOR FILING DATE: 1995-04-12
 / PRIOR APPLICATION NUMBER: 08/346,731
 / PRIOR FILING DATE: 1994-11-21
 / NUMBER OF SEQ ID NOS: 12485
 / SOFTWARE: PatentIn Ver. 2.0
 / SEQ ID NO 6019
 / LENGTH: 490
 / TYPE: DNA
 / ORGANISM: Homo sapiens
 / FEATURE:
 / NAME/KEY: misc feature
 / LOCATION: (10)
 / OTHER INFORMATION: n equals a,t,g, or c
 / NAME/KEY: misc feature
 / LOCATION: (58)
 / OTHER INFORMATION: n equals a,t,g, or c
 / NAME/KEY: misc feature
 / LOCATION: (119)
 / OTHER INFORMATION: n equals a,t,g, or c
 / NAME/KEY: misc feature
 / LOCATION: (134)
 / OTHER INFORMATION: n equals a,t,g, or c
 / NAME/KEY: misc feature
 / LOCATION: (140)
 / OTHER INFORMATION: n equals a,t,g, or c
 / NAME/KEY: misc feature
 / LOCATION: (162)
 / OTHER INFORMATION: n equals a,t,g, or c
 / NAME/KEY: misc feature
 / LOCATION: (203)
 / OTHER INFORMATION: n equals a,t,g, or c
 / NAME/KEY: misc feature
 / LOCATION: (205)
 / OTHER INFORMATION: n equals a,t,g, or c
 / NAME/KEY: misc feature
 / LOCATION: (262)
 / OTHER INFORMATION: n equals a,t,g, or c
 / NAME/KEY: misc feature
 / LOCATION: (266)
 / OTHER INFORMATION: n equals a,t,g, or c
 / NAME/KEY: misc feature
 / LOCATION: (285)
 / OTHER INFORMATION: n equals a,t,g, or c
 / NAME/KEY: misc feature
 / LOCATION: (294)
 / OTHER INFORMATION: n equals a,t,g, or c
 / NAME/KEY: misc feature
 / LOCATION: (299)
 / OTHER INFORMATION: n equals a,t,g, or c
 / NAME/KEY: misc feature

/ LOCATION: (301)
 / OTHER INFORMATION: n equals a,t,g, or c
 / NAME/KEY: misc feature
 / LOCATION: (305)
 / OTHER INFORMATION: n equals a,t,g, or c
 / NAME/KEY: misc feature
 / LOCATION: (322)
 / OTHER INFORMATION: n equals a,t,g, or c
 / NAME/KEY: misc feature
 / LOCATION: (327)
 / OTHER INFORMATION: n equals a,t,g, or c
 / NAME/KEY: misc feature
 / LOCATION: (347)
 / OTHER INFORMATION: n equals a,t,g, or c
 / NAME/KEY: misc feature
 / LOCATION: (358)
 / OTHER INFORMATION: n equals a,t,g, or c
 / NAME/KEY: misc feature
 / LOCATION: (365)
 / OTHER INFORMATION: n equals a,t,g, or c
 / NAME/KEY: misc feature
 / LOCATION: (370)
 / OTHER INFORMATION: n equals a,t,g, or c
 / NAME/KEY: misc feature
 / LOCATION: (372)
 / OTHER INFORMATION: n equals a,t,g, or c
 / NAME/KEY: misc feature
 / LOCATION: (376)
 / OTHER INFORMATION: n equals a,t,g, or c
 / NAME/KEY: misc feature
 / LOCATION: (388)
 / OTHER INFORMATION: n equals a,t,g, or c
 / NAME/KEY: misc feature
 / LOCATION: (397)
 / OTHER INFORMATION: n equals a,t,g, or c
 / NAME/KEY: misc feature
 / LOCATION: (401)
 / OTHER INFORMATION: n equals a,t,g, or c
 / NAME/KEY: misc feature
 / LOCATION: (403)
 / OTHER INFORMATION: n equals a,t,g, or c
 / NAME/KEY: misc feature
 / LOCATION: (424)
 / OTHER INFORMATION: n equals a,t,g, or c
 / NAME/KEY: misc feature
 / LOCATION: (426)
 / OTHER INFORMATION: n equals a,t,g, or c
 / NAME/KEY: misc feature
 / LOCATION: (427)
 / OTHER INFORMATION: n equals a,t,g, or c
 / NAME/KEY: misc feature
 / LOCATION: (430)
 / OTHER INFORMATION: n equals a,t,g, or c
 / NAME/KEY: misc feature
 / LOCATION: (431)
 / OTHER INFORMATION: n equals a,t,g, or c
 / NAME/KEY: misc feature
 / LOCATION: (433)
 / OTHER INFORMATION: n equals a,t,g, or c
 / NAME/KEY: misc feature
 / LOCATION: (434)
 / OTHER INFORMATION: n equals a,t,g, or c
 / NAME/KEY: misc feature
 / LOCATION: (449)
 / OTHER INFORMATION: n equals a,t,g, or c
 / NAME/KEY: misc feature
 / LOCATION: (450)
 / OTHER INFORMATION: n equals a,t,g, or c
 / NAME/KEY: misc feature
 / LOCATION: (452)
 / OTHER INFORMATION: n equals a,t,g, or c
 / NAME/KEY: misc feature
 / LOCATION: (461)

```

? OTHER INFORMATION: n equals a,t,g, or c
? NAME/KEY: misc feature
? LOCATION: (462)
? OTHER INFORMATION: n equals a,t,g, or c
? NAME/KEY: misc feature
? LOCATION: (467)
? OTHER INFORMATION: n equals a,t,g, or c
? NAME/KEY: misc feature
? LOCATION: (472)
? OTHER INFORMATION: n equals a,t,g, or c
? NAME/KEY: misc feature
? LOCATION: (473)
? OTHER INFORMATION: n equals a,t,g, or c
? NAME/KEY: misc feature
? LOCATION: (476)
? OTHER INFORMATION: n equals a,t,g, or c
? NAME/KEY: misc feature
? LOCATION: (489)
? OTHER INFORMATION: n equals a,t,g, or c
? 590-783-590-6019

```

Query Match	3.6%	Score 38.2	DB 10	Length 490
Best Local Similarity	55.1%	Pred. No. 2.6		
Matches 70	Conservative 0	Mismatches 57	Indels 0	Gaps 0

OY 739 AATGTTCACTCATTTGGTACCTCTGTATAAGCCATCATTATACTGATGTTTGCGTTTTG 798
| | | | | | | | | | | | | | | | | | | | |
Db 69 AGTATGCAGAGCTTAGGAATCACTGGATPACAAAAAGGTTATATGGAAGTNTTCCTTTTG 128

OY 799 AATTCACATTGTGTAAGAAAAGCCTTGGTGGAAGTTATCAAGATTATTTG 858
| | | | | | | | | | | | | | |
Db 129 CATGNAGTTNTGAACGAATGCACCTTNTGTGCAATTAATCCAGATTCTTGA 188

QY	859	ACTAATA	865
Db	189	AGGAATA	195

RESULT 6

```

US-09-834-975-317/c
; Sequence 317, Application US/09834975
; Patent No. US20020110815A1
; GENERAL INFORMATION:
; APPLICANT: Lillie, James
; APPLICANT: Brown, Jeffrey
; APPLICANT: Bolt, Andrew
; APPLICANT: Van Hufel, Christophe
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS AND METHODS
; TITLE OF INVENTION: FOR THE IDENTIFICATION, ASSESSMENT, PREVENTION, AND THERAPY
; TITLE OF INVENTION: OF HUMAN CANCERS
; FILE REFERENCE: MRI-016B
; CURRENT APPLICATION NUMBER: US/09/834,975
; CURRENT FILING DATE: 2001-04-13
; PRIOR APPLICATION NUMBER: 60/197,538
; PRIOR FILING DATE: 2000-04-14
; NUMBER OF SEQ ID NOS: 1046
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 317
; LENGTH: 536
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1) - (536)
; OTHER INFORMATION: n = A,T,C OR G
US-09-834-975-317

```

	Query Match	Similarity	Score	DB	length
	Best Local	46.8%	Pred. No. 3.8	107	Indels 0; Gaps 0;
	Matches	94	Conservative		
0y	334	GGTTTCCCAACATATCTTGCTCAAACTTTTATTATGCTCTACCAAAATGTACTACAA	393		

Accession	Sequence	Position
Db	366 GTTTTNCCTTTTTCNTTTTNNAAAAANNTTTTITTTTAAAAAANAATTCGNAAT	307
Oy	394 GCATCTTCCTTTTGAAGATTGGCTCAAAABAGTTGGTATGACAAATTTGTTATCCAGAT	453
Db	306 TTCCCCCGCCTTTTAAAAACGGCCCAAAAAAGATNTTTAAAAAGTTTGGGCNCAAAA	247
Oy	454 GTTGCCCTTCTAATGTTTTCACAGTTTGCGATTGATCAACAGATTAATCCAAATCATCA	513
Db	246 AAAAAAGCCTTTCCTTNTATAAAGTTTGTATTATTTTAAAGCAATGTTNCCATTGACA	187
Oy	514 TCATCAGTCGAGTTGATATTC	534
Db	186 NCAAAATTCAGATGCTAGATC	166

RESULT 7

US-09-070-927A-180/c
; Sequence 180, Application US/09070927A
; Patent No. US20020120116A1
; GENERAL INFORMATION:
; Information about the work

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

CITY: Rockville
 STATE: Maryland
 COUNTRY: USA

COMPUTER READABLE FORM ;

```

MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
COMPUTER: HP Vectra 486/33

```

```
;
;
; OPERATING SYSTEM: MSDOS version 6.22
; SOFTWARE: ASCII Text
```

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/070,927A

FILING DATE: 04-May-2000
CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/046,655

APPLICATION NUMBER: 60/044,031
FILING DATE: 1997-05-16
APPLICATION NUMBER: 60/044,031

APPLICATION NUMBER: 60/066,009
FILING DATE: 1997-05-06
APPLICATION NUMBER: 60/066,009

APPLICATION NUMBER: 80/088,000
FILING DATE: 1997-11-14
ATTORNEY/AGENT INFORMATION:

ATTORNEY/AGENT INFORMATION:
NAME: Kenley K. Hoover
REGISTRATION NUMBER: 40 303

REGISTRATION NUMBER: 40,302
REFERENCE/DOCKET NUMBER: PB369
RECONSTRUCTION INFORMATION:

TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
FAX: (301) 300-8713

TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 180:
; INFORMATION FOR SEQ ID NO: 180:
; INFORMATION FOR SEQ ID NO: 180:

```

;
; SEQUENCE CHARACTERISTICS:
; LENGTH: 31517 base pairs
;

```

```

; TYPE: nucleic acid
; STRANDEDNESS: double

```

```

; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 18

```

US-09-070-927A-180

Query Match	3.5%	Score 37.2	DB 10	Length 31517
Best Local Similarly	55.4%	Pred. No. 20		
Matches 72; Conservative	0	Mismatches 58;	Indels 0;	Gaps 0;

QY 189 TGTTTTCTCAACTGATACCATTTTCCCTTCATATCTTGTTCTTTGGGTGGGTCAAG 248
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 24369 TGTTCTTCATCATTAAGTTATCCTTGTCGTGGTTGTTACCTTGGGCTTTTTCGC 24310

QY 249 TTTCGATGAATCATTTGGTTCGTCGCGTTACTGCGATTTGCATAGTTTTGTTTCTCAAG 308


```
Db 24309 TGTATATGATGATGTTGCTTCTTACTTGTGTTTGTACTGTAATGTTTCTGCAA 24250
QY 309 TGCCCTGTGC 318
Db 24249 AACAGTTGTC 24240

RESULT 8
US-09-946-807-1
; Sequence 1, Application US/09946807
; Patent No. US20020165144A1
; GENERAL INFORMATION:
; APPLICANT: Stefansson, Hreinn
; APPLICANT: Steinhorsdotcitr, Valgerdur
; APPLICANT: Gulcher, Jeffrey R.
; TITLE OF INVENTION: HUMAN SCHIZOPHRENIA GENE
; FILE REFERENCE: 2345.2004-001
; CURRENT APPLICATION NUMBER: US/09/946,807
; PRIOR FILING DATE: 2001-09-05
; PRIOR APPLICATION NUMBER: US/09/795,668
; PRIOR FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: US 09/515,716
; PRIOR FILING DATE: 2000-02-28
; NUMBER OF SEQ ID NOS: 1531
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1503841
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(1531)
; OTHER INFORMATION: m=a or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(1531)
; OTHER INFORMATION: k=g or t/u
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(1531)
; OTHER INFORMATION: s=g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(1531)
; OTHER INFORMATION: w=a or t/u
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(1531)
; OTHER INFORMATION: b=g or c or t/u
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(1531)
; OTHER INFORMATION: d=a or g or t/u
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(1531)
; OTHER INFORMATION: h=a or c or t/u
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(1531)
; OTHER INFORMATION: v=a or g or c
; NAME/KEY: misc_feature
; LOCATION: (1)...(1531)
```

```
; OTHER INFORMATION: n=a or g or c or t/u
US-09-946-807-1
Query Match 3.4%; Score 36.6; DB 9; Length 1503841;
Best Local Similarity 50.9%; Pred. No. 93;
Matches 87; Conservative 0; Mismatches 84; Indels 0; Gaps 0;

QY 319 AATGTTCCGTTATAGTTATGTTCCCAACAATCTGTGCTCAAACTTTTATGTGCTCACA 378
Db 1028986 AATTTCTACAAAGTTTATGTCTCTGCGCCAGTCTTAAGCCCTTTCTAGATAACATCCA 1029045

QY 379 AATGTTCTCAACAAGATCTTCCCTTTTGAATTTGGCTTAAAGTTGGTTATGAACAA 438
Db 1029046 GAAGATTAATCATCCACAGACTATCCGGGTACATTCGGCTGACTTGGTTAGTGGCTTAA 1029105

QY 439 AATGTTATCCAGATGTTGCGCTTCTAGTCTTTGCACAGTTTTCGGATTG 489
Db 1029106 GCAGTATATCCATGACTTAACTGCGGTTTCTCAAACTTTTTCCTTTG 1029156

RESULT 9
US-09-795-668-1
; Sequence 1, Application US/09795668
; Patent No. US20020045577A1
; GENERAL INFORMATION:
; APPLICANT: Stefansson, Hreinn
; APPLICANT: Steinhorsdotcitr, Valgerdur
; APPLICANT: Gulcher, Jeffrey R.
; TITLE OF INVENTION: HUMAN SCHIZOPHRENIA GENE
; FILE REFERENCE: 2345.2004-001
; CURRENT APPLICATION NUMBER: US/09/795,668
; PRIOR FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: US 09/515,716
; PRIOR FILING DATE: 2000-02-28
; NUMBER OF SEQ ID NOS: 1531
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1503841
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(1531)
; OTHER INFORMATION: r=g or a
; NAME/KEY: misc_feature
; LOCATION: (1)...(1531)
; OTHER INFORMATION: y=c/u or c
; NAME/KEY: misc_feature
; LOCATION: (1)...(1531)
; OTHER INFORMATION: m=a or c
; NAME/KEY: misc_feature
; LOCATION: (1)...(1531)
; OTHER INFORMATION: k=g or t/u
; NAME/KEY: misc_feature
; LOCATION: (1)...(1531)
; OTHER INFORMATION: s=g or c
; NAME/KEY: misc_feature
; LOCATION: (1)...(1531)
; OTHER INFORMATION: w=a or t/u
; NAME/KEY: misc_feature
; LOCATION: (1)...(1531)
; OTHER INFORMATION: b=g or c or t/u
; NAME/KEY: misc_feature
; LOCATION: (1)...(1531)
; OTHER INFORMATION: d=a or g or t/u
; NAME/KEY: misc_feature
; LOCATION: (1)...(1531)
; OTHER INFORMATION: h=a or c or t/u
; NAME/KEY: misc_feature
; LOCATION: (1)...(1531)
; OTHER INFORMATION: v=a or g or c
; NAME/KEY: misc_feature
; LOCATION: (1)...(1531)
```

OTHER INFORMATION: n=a or g or c or t/u
US-09-795-668-1

Query Match 3.4%; Score 36.6; DB 10; Length 1503841;
Best Local Similarity 50.9%; Pred. No. 93;
Matches 87; Conservative 0; Mismatches 84; Indels 0; Gaps 0;

QY 319 AATGGTCCGTTATGGTTCCCAATACCTTGTGCTCAACCTTTATATGTCCTACCA 378
DB 1028986 AATTCTACAAGTTATGTCTCTGCCAAGTTCTTAAGCCCTCTTCTAGATACCATCCA 1029045

QY 379 AATGTTCTCAACAGCATCTTCCCTTTTGAATGGCTCAAAAGTTGGTTAGAACAA 438
DB 1029046 GAAGATAATCATCCACAGCATTCGGGTACATTCGGCTGACTGTAGTTGGGTCTAA 1029105

QY 439 AATGTTATCCAGATGTTGCCTTTAGTCTTTCACAGTTTCGATG 489
DB 1029106 GCAGTATATCATGATGTTCTAATCTCGGGTTCTCAAACTTTTCTCTTG 1029156

RESULT 10
US-09-795-686-1

Sequence 1, Application US/09795686
Patent No. US20020094954A1

GENERAL INFORMATION:

APPLICANT: Stefansson, Hreinn

APPLICANT: Steinhorsdottir, Valgerdur

APPLICANT: Gulcher, Jeffrey R.

APPLICANT: Zhu, Tong

TITLE OF INVENTION: HUMAN SCHIZOPHRENIA GENE

FILE REFERENCE: 2345.2005-001

CURRENT APPLICATION NUMBER: US/09/795,686

PRIOR FILING DATE: 2001-02-28

PRIOR APPLICATION NUMBER: US 09/515,715

PRIOR FILING DATE: 2000-02-28

NUMBER OF SEQ ID NOS: 1531

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 1

LENGTH: 1503841

TYPE: DNA

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: misc_feature

LOCATION: (1)..(1531)

OTHER INFORMATION: r=g or a

NAME/KEY: misc_feature

LOCATION: (1)..(1531)

OTHER INFORMATION: y=c/u or c

NAME/KEY: misc_feature

LOCATION: (1)..(1531)

OTHER INFORMATION: m=a or c

NAME/KEY: misc_feature

LOCATION: (1)..(1531)

OTHER INFORMATION: k=g or t/u

NAME/KEY: misc_feature

LOCATION: (1)..(1531)

OTHER INFORMATION: s=g or c

NAME/KEY: misc_feature

LOCATION: (1)..(1531)

OTHER INFORMATION: w=a or t/u

NAME/KEY: misc_feature

LOCATION: (1)..(1531)

OTHER INFORMATION: d=a or g or t/u

NAME/KEY: misc_feature

LOCATION: (1)..(1531)

OTHER INFORMATION: b=a or c or t/u

NAME/KEY: misc_feature

LOCATION: (1)..(1531)

OTHER INFORMATION: v=a or g or c

NAME/KEY: misc_feature

LOCATION: (1)..(1531)

OTHER INFORMATION: n=a or g or c or t/u
US-09-795-686-1

Query Match 3.4%; Score 36.6; DB 10; Length 1503841;
Best Local Similarity 50.9%; Pred. No. 93;
Matches 87; Conservative 0; Mismatches 84; Indels 0; Gaps 0;

QY 319 AATGGTCCGTTATGGTTCCCAATACCTTGTGCTCAACCTTTATATGTCCTACCA 378
DB 1028986 AATTCTACAAGTTATGTCTCTGCCAAGTTCTTAAGCCCTCTTCTAGATACCATCCA 1029045

QY 379 AATGTTCTCAACAGCATCTTCCCTTTTGAATGGCTCAAAAGTTGGTTAGAACAA 438
DB 1029046 GAAGATAATCATCCACAGCATTCGGGTACATTCGGCTGACTGTAGTTGGGTCTAA 1029105

QY 439 AATGTTATCCAGATGTTGCCTTTAGTCTTTCACAGTTTCGATG 489
DB 1029106 GCAGTATATCATGATGTTCTAATCTCGGGTTCTCAAACTTTTCTCTTG 1029156

RESULT 11
US-09-938-842A-4858/C

Sequence 4858, Application US/09938842A
Patent No. US20020160378A1

GENERAL INFORMATION:

APPLICANT: Harper, Jeff

APPLICANT: Kreps, Joel

APPLICANT: Wang, Xun

APPLICANT: Zhu, Tong

TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING

FILE REFERENCE: SAME, AND METHODS OF USE

CURRENT APPLICATION NUMBER: US/09/938,842A

PRIOR FILING DATE: 2001-08-24

PRIOR APPLICATION NUMBER: US 60/227,866

PRIOR FILING DATE: 2000-08-24

PRIOR APPLICATION NUMBER: US 60/264,647

PRIOR FILING DATE: 2001-01-16

PRIOR APPLICATION NUMBER: US 60/300,111

PRIOR FILING DATE: 2001-06-22

NUMBER OF SEQ ID NOS: 5379

SEQ ID NO 4858

LENGTH: 2000

TYPE: DNA

ORGANISM: Arabidopsis thaliana

US-09-938-842A-4858

Query Match 3.4%; Score 35.8; DB 9; Length 2000;
Best Local Similarity 48.7%; Pred. No. 17;
Matches 97; Conservative 0; Mismatches 102; Indels 0; Gaps 0;

QY 669 TTCAACTACTANGTCGGATATAGTGAAGATGTGAATTAAGATATCATCAGAGA 728
DB 1059 TTCACTTGTGGTCAATATACAGTAACATTTATTCAGGGTGACATTAACGATA 1000

QY 729 TCAACAATATACATGTTCAACTCATTTGTAAGCTGTATAGCATATGATGATGT 788
DB 999 TCCACACACAGACAGATTTATTAAGATTGTATTTAGCCCTGATATTTGGTTCAA 940

QY 789 TTGGCTTTGAATTCATTTGTGTGATGAAGAACAAAGGTTGCTGTGAAGTTATCAA 848
DB 939 AATATACCTTAATATACATTTCAATTTCTTTAAATAAATTTATTTGAATAGAGAACAA 880

QY 849 GAATTTATGACTAATCT 867
DB 879 AATTTACCGAAAAATATT 861

RESULT 12

US-09-818-512-3

Sequence 3, Application US/09818512

Patent No. US20020142416A1

GENERAL INFORMATION:

APPLICANT: BEASLEY, Ellen et al.
 TITLE OF INVENTION: ISOLATED HUMAN ENZYME PROTEINS, NUCLEIC
 TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN ENZYME PROTEINS, AND USES
 TITLE OF INVENTION: THEREOF
 FILE REFERENCE: CL001192
 CURRENT APPLICATION NUMBER: US/09/818,512
 CURRENT FILING DATE: 2001-03-28
 NUMBER OF SEQ ID NOS: 4
 SOFTWARE: FastSeq for Windows Version 4.0
 SEQ ID NO 3
 LENGTH: 116592
 TYPE: DNA
 ORGANISM: Human
 FEATURE:
 NAME/KEY: misc feature
 LOCATION: (1)...(116592)
 OTHER INFORMATION: n = A,T,C or G
 US-09-818-512-3

Query Match 3.4%; Score 35.8; DB 10; Length 116592;
 Local Similarity 52.3%; Pred. No. 70;
 Matches 79; Conservative 0; Mismatches 72; Indels 0; Gaps 0;

QY 706 GAATTAGGATATCATGAGATCAACATATCAATGTTCACTGATCTTGAT 765
 DB 13526 GATATAGCTATATCATGAGATCAACATATCAATGTTCACTGATCTTGAT 13585
 QY 766 AAGCATACGTTTACTGATGTTTGGCTTGAATTCATTTGTGAGAAAGCA 825
 DB 13586 TATTATATCATGTTTACTGATGATGAAATTTGATTTGTTGAGCTAAGGAC 13645
 QY 826 AAGTGTCTGTGAAGTTATCAAGATTTAT 856
 DB 13646 ATTTTGAAGGTGTTCTCTTCAAGATTTCT 13676

RESULT 13
 US-09-815-242-3894/c
 Sequence 3894, Application US/09815242
 Patent No. US20020061569A1
 GENERAL INFORMATION:
 APPLICANT: Haselbeck, Robert
 APPLICANT: Ohlsen, Karl L.
 APPLICANT: Zyskind, Judith W.
 APPLICANT: Wall, Daniel
 APPLICANT: Trawick, John D.
 APPLICANT: Carr, Grant J.
 APPLICANT: Yamamoto, Robert T.
 TITLE OF INVENTION: Identification of Essential Genes in
 TITLE OF INVENTION: Prokaryotes
 FILE REFERENCE: ELITRA.011A
 CURRENT APPLICATION NUMBER: US/09/815,242
 CURRENT FILING DATE: 2001-03-21
 PRIOR APPLICATION NUMBER: 60/191,078
 PRIOR FILING DATE: 2000-03-21
 PRIOR APPLICATION NUMBER: 60/206,848
 PRIOR FILING DATE: 2000-05-23
 PRIOR APPLICATION NUMBER: 60/207,727
 PRIOR FILING DATE: 2000-05-26
 PRIOR APPLICATION NUMBER: 60/242,578
 PRIOR FILING DATE: 2000-10-23
 PRIOR APPLICATION NUMBER: 60/253,625
 PRIOR FILING DATE: 2000-11-27
 PRIOR APPLICATION NUMBER: 60/257,931
 PRIOR FILING DATE: 2000-12-22
 PRIOR APPLICATION NUMBER: 60/269,308
 PRIOR FILING DATE: 2001-02-16
 NUMBER OF SEQ ID NOS: 14110
 SOFTWARE: FastSeq for Windows Version 4.0
 SEQ ID NO 3894
 LENGTH: 2295
 TYPE: DNA

ORGANISM: Enterococcus faecalis
 US-09-815-242-3894

Query Match 3.3%; Score 35.6; DB 10; Length 2295;
 Best Local Similarity 54.6%; Pred. No. 20;
 Matches 71; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

QY 189 TGTTCCTCACTGATGATGATTTTCTTCATATCTTGTTCCTTGGGTGTCAGAG 248
 DB 204 TGTTCCTCACTGATGATGATTTTCTTCATATCTTGTTCCTTGGGTGTCAG 145
 QY 249 TTGGATGAATCATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 308
 DB 144 TGTTAATGAATGATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 85
 QY 309 TGGCTCTGTC 318
 DB 84 AACAGTTGTC 75

RESULT 14
 US-09-815-242-6496/c
 Sequence 6496, Application US/09815242
 Patent No. US20020061569A1
 GENERAL INFORMATION:
 APPLICANT: Haselbeck, Robert
 APPLICANT: Ohlsen, Karl L.
 APPLICANT: Zyskind, Judith W.
 APPLICANT: Wall, Daniel
 APPLICANT: Trawick, John D.
 APPLICANT: Carr, Grant J.
 APPLICANT: Yamamoto, Robert T.
 APPLICANT: Xu, H. Howard
 TITLE OF INVENTION: Identification of Essential Genes in
 TITLE OF INVENTION: Prokaryotes
 FILE REFERENCE: ELITRA.011A
 CURRENT APPLICATION NUMBER: US/09/815,242
 CURRENT FILING DATE: 2001-03-21
 PRIOR APPLICATION NUMBER: 60/191,078
 PRIOR FILING DATE: 2000-03-21
 PRIOR APPLICATION NUMBER: 60/206,848
 PRIOR FILING DATE: 2000-05-23
 PRIOR APPLICATION NUMBER: 60/207,727
 PRIOR FILING DATE: 2000-05-26
 PRIOR APPLICATION NUMBER: 60/242,578
 PRIOR FILING DATE: 2000-10-23
 PRIOR APPLICATION NUMBER: 60/253,625
 PRIOR FILING DATE: 2000-11-27
 PRIOR APPLICATION NUMBER: 60/257,931
 PRIOR FILING DATE: 2000-12-22
 PRIOR APPLICATION NUMBER: 60/269,308
 PRIOR FILING DATE: 2001-02-16
 NUMBER OF SEQ ID NOS: 14110
 SOFTWARE: FastSeq for Windows Version 4.0
 SEQ ID NO 6496
 LENGTH: 2295
 TYPE: DNA
 ORGANISM: Enterococcus faecalis
 FEATURE:
 NAME/KEY: CDS
 LOCATION: (1)...(2295)
 US-09-815-242-6496

Query Match 3.3%; Score 35.6; DB 10; Length 2295;
 Best Local Similarity 54.6%; Pred. No. 20;
 Matches 71; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

QY 189 TGTTCCTCACTGATGATGATTTTCTTCATATCTTGTTCCTTGGGTGTCAGAG 248
 DB 204 TGTTCCTCACTGATGATGATTTTCTTCATATCTTGTTCCTTGGGTGTCAG 145
 QY 249 TTGGATGAATCATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 308

Db 144 TGTAAATGAATGATGTCCTCTCTACTTGTGTTGACTGTGTAATGTTCTGCAA 85
 QY 309 TGCCTGTGC 318
 Db 84 AACAGTTGTC 75

RESULT 15

US-09-796-692-2970
 ; Sequence 2970, Application US/09796692
 ; Publication No. US20020198362A1

GENERAL INFORMATION:
 ; APPLICANT: Gaiger, Alexander

APPLICANT: Algate, Paul A.

APPLICANT: Mannion, Jane

TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THERAPY

FILE REFERENCE: 2077,001200

CURRENT FILING DATE: 2001-03-01

PRIOR FILING DATE: 2000-03-01

PRIOR FILING DATE: 2000-03-01

PRIOR FILING DATE: 2000-03-17

PRIOR FILING DATE: 2000-03-17

PRIOR FILING DATE: 2000-04-27

PRIOR FILING DATE: 2000-04-27

PRIOR FILING DATE: 2000-04-28

PRIOR FILING DATE: 2000-04-28

PRIOR FILING DATE: 2000-04-28

PRIOR FILING DATE: 2000-05-01

PRIOR FILING DATE: 2000-05-01

PRIOR FILING DATE: 2000-05-04

PRIOR FILING DATE: 2000-05-04

PRIOR FILING DATE: 2000-05-22

PRIOR FILING DATE: 2000-07-14

PRIOR FILING DATE: 2000-07-14

PRIOR FILING DATE: 2000-08-03

PRIOR FILING DATE: 2000-08-03

PRIOR FILING DATE: 2000-08-04

PRIOR FILING DATE: 2000-08-07

NUMBER OF SEQ ID NOS: 9597

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 2970

LENGTH: 371

TYPE: DNA

ORGANISM: Homo sapiens

US-09-796-692-2970

Query Match 3.3%; Score 35.2; DB 9; Length 371;
 Best Local Similarity 50.6%; Pred. No. 13;

Matches 85; Conservative 0; Mismatches 83; Indels 0; Gaps 0;

QY 582 CACCTCAAAATATATGAACTTTTGGATTCAACAGTTGTCCTCAAGGAATATAT 641
 Db 129 CACTGGAACCAACCAATTCAGGTCGAAATCGTATTAATGACACAAACATTTA 188
 QY 642 TAACTGTGTAAGCAAGGTAACCAATTTCAACTACTATGCGATATAGTGAAGTAT 701
 Db 189 CAGTTGTCAGACAGATTCATTTTCATTAACCAATTAACAAATTAACAAATTTTAT 248
 QY 702 GTGTGAATTAGATATCATCAGATCAACAATATACATGTTCACT 749
 Db 249 GTGTAGATTAATTAATCTTTGACTGCTGAAAAATTAATTTTAAAT 296

Search completed: January 27, 2003, 07:48:41
 Job time : 4938 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Comugen Ltd.

OM nucleic - nucleic search, using SW model

Run on: January 27, 2003, 04:02:24 ; Search time 1718 Seconds
(without alignments)
10067.974 Million cell updates/sec

Title: US-09-675-509-3
Perfect score: 1068
Sequence: 1 atgtccaccacacaaagac.....gtgtcttgagaccatccct 1068

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues
T number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST.*
1: em_gesbda:*
2: em_gesbhm:*
3: em_geslin:*
4: em_geslun:*
5: em_geslov:*
6: em_geslpi:*
7: em_geslro:*
8: em_geslro:*
9: gb_gesl1:*
10: gb_gesl2:*
11: gb_gesl3:*
12: gb_gesl4:*
13: gb_gesl5:*
14: gb_gesl6:*
15: em_geslun:*
16: em_geslun:*
17: gb_gesl7:*
18: em_geslun:*
19: em_geslun:*
20: em_geslun:*
21: em_geslun:*
22: em_geslun:*
23: em_geslun:*
24: em_geslun:*
25: em_geslun:*
26: em_geslun:*
27: em_geslun:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	51.8	4.9	1101	17	CNS000001
2	51.8	4.9	1225	17	CNS00161D
3	44.6	4.2	1101	17	CNS00038G
4	44.2	4.1	887	17	CNS00005M
5	44	4.1	518	12	BGS53567
6	44	4.1	990	17	CNS00060I

C	7	44	4.1	1101	17	CNS000EVL	AL069706 Drosophila
C	8	44	4.1	1101	17	CNS001219	AL101595 Drosophila
C	9	43.8	4.1	520	10	AM774971	AM774971 EST34122
C	10	43.6	4.1	537	10	BM276144	BM276144 PEST0A87
C	11	43.6	4.1	530	13	BM273621	BM273621 PEST0A85
C	12	43.6	4.1	533	13	BM276328	BM276328 PEST0A88
C	13	43.6	4.1	570	13	BM275241	BM275241 PEST0A86
C	14	43.6	4.1	747	17	CNS0011RO	AL100640 Drosophila
C	15	43.6	4.1	1101	17	CNS00039G	AL063921 Drosophila
C	16	43.6	4.1	1101	17	CNS00022U	AL097152 Drosophila
C	17	43.4	4.1	649	12	BF298250	BF298250 015BDG01
C	18	43.4	4.1	870	17	CNS0009XJ	AL054462 Drosophila
C	19	43.4	4.1	1001	17	CNS00155H	AL105023 Drosophila
C	20	43.4	4.1	1006	17	CNS0010EY	AL098884 Drosophila
C	21	43.4	4.0	521	17	AZ336424	AZ336424 1M0066M10
C	22	43.4	4.0	628	17	AZ818307	AZ818307 2M0088R14
C	23	43.4	4.0	701	17	AZ316576	AZ316576 1M0034H16
C	24	42.8	4.0	467	17	CNS0011DW	AL100142 Drosophila
C	25	42.8	4.0	526	13	B1814808	B1814808 PEST0A80
C	26	42.8	4.0	668	9	AL514901	AL514901 AL514901
C	27	42.8	4.0	734	17	CNS0010MP	AL099163 Drosophila
C	28	42.8	4.0	869	17	CNS0017Z2	AL108680 Drosophila
C	29	42.8	4.0	1121	9	AL554111	AL554111 AL554111
C	30	42.6	4.0	521	17	AQ810997	AQ810997 HS 5382_A
C	31	42.6	4.0	780	17	AQ842859	AQ842859 CP31294B
C	32	42.6	4.0	1086	17	CNS012XX	AL102146 Drosophila
C	33	42.4	4.0	557	17	AOS05170	AOS05170 RPCR-11-2
C	34	42.4	4.0	978	17	AG129148	AG129148 Pan tlog1
C	35	42.4	4.0	996	17	CNS000FUH	AL071063 Drosophila
C	36	42.4	4.0	1101	17	CNS0012TP	AL102007 Drosophila
C	37	42.2	4.0	409	17	AQ455933	AQ455933 HS 5151_B
C	38	42.2	4.0	658	17	B67167	B67167 CpG0046A_Cp
C	39	42.2	4.0	1101	17	CNS017KX	AL108171 Drosophila
C	40	42.2	4.0	1191	14	BQ421837	BQ421837 AGENCOURT
C	41	42.2	3.9	839	12	BG400936	BG400936 602464975
C	42	42	3.9	1004	9	AZ136058	AL551881 AL551881
C	43	41.6	3.9	803	17	AZ136058	SP 0171_A
C	44	41.6	3.9	811	17	CNS003YOO	AL266649 Tetraodon
C	45	41.6	3.9	963	17	CNS000A4L	AL054918 Drosophila

ALIGNMENTS

RESULT 1
CNS000001/c 1101 bp DNA linear GSS 03-JUN-1999
LOCUS Drosophila melanogaster genome survey sequence TET3 end of BAC #
DEFINITION BACR01J16 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.

ACCESSION AL065414.1 GI:4938827
VERSION GSS.
KEYWORDS Drosophila melanogaster.
SOURCE Drosophila melanogaster
ORGANISM Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.

REFERENCE 1 (bases 1 to 1101)
Genoscope.

AUTHORS Direct Submission
TITLE Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
JOURNAL BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)

COMMENT

Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osogawa and Aaron Mamoser in Pletier de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial

SOURCE	ORGANISM	REFERENCE	AUTHORS	COMMENT
Source	Drosophila melanogaster	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pezomyzeta; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.	1 (bases 1 to 1101)	Genoscope.
Journal	Submitted (02-JUN-1999)	Genoscope - Centre National de Sequencage	BP 191 91006 Evry cedex - FRANCE (E-mail : sequei@genoscope.cns.fr - Web : www.genoscope.cns.fr)	Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Oosagawa and Aaron Mamoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain Y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm .
FEATURES	source	1..1101	/organism="Drosophila melanogaster"	
			/db_xref="taxon:7227"	
			/clone="BACR08K10"	
			/clone_11b="RPI-98"	
			/note="end : TET3"	
BASE COUNT	201 a	64 c	131 g	202 t 503 others
ORIGIN				
Query Match	4.2%	Score 44.6;	DB 17;	Length 1101;
Best Local Similarity	21.7%	Pred. No. 1.5;		
Matches	89;	Conservative 139;	Mismatches 183;	Indels 0; Gaps 0;
593	ATAGAACCATTTGGATTCACAGTGTGCTGCCTCCAAAGAAATATNTAATCTGTA	652		
805	RKRPRKAAADRRDAADDDNAATTTTTRTDIDDMKMTDTWTRAAADRTDRDD	746		
653	AGCAAGTAAACCAATTTCAACTACTATGTCGATATGTAAGTATGTGAAATTA	712		
745	DDRDAAGTAAGRKRTWKRMRKRTDADADTARDRRRRGGDADAKGKTKRK	686		
713	AGGATATCATCAAGATCAACATCAATGTTCAACTATTGTA	772		
685	RRRRDRATWDRDADAAAMWTTTDTDDDKDRRKGARRRRRTTARAAAMWMTK	626		
773	ACGTTATACGATGTTTGGCTTGGATTTCCAAATTTGTCATGAAAAGCAAGTTG	832		
625	AMDWAKMWMKTRPADWDRAADTWDARADDMKARMRARRRARAARDRTWTKG	566		
833	CTGTGAAGTTATCAAGAAATTTATGACAACTAGTTGCTGGACTGTGGCTCG	892		
565	KTTTATTTWTAARAAMWMAATTTATTTTWTWTWTWTWTWTWTWTWTWTWT	506		
893	GATTACTCTCCGACCAAGAAATGATTTGCTCATTTGGCTAAATCATCAACTTT	952		
505	WTATTAATTAAMWMAAATTTTWTWTWTWTWTWTWTWTWTWTWTWTWTWT	446		
953	ATGCTCAATTTGAGCAACAAATTCATGCCAAGAAAGTGAAGTT	1003		
445	AATTTTWTWTWTAAATTTTWTWTWTWTWTWTWTWTWTWTWTATATK	395		
RESULT 4	CNS00D5M	887 bp	DNA	linear
LOCUS	CNS00D5M			
DEFINITION	Drosophila melanogaster genome survey sequence TET3 end of BAC #			

ACCESSION	BACR26H11	of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
VERSION	AL060256	
KEYWORDS	AL060256.1 GI:4947618 GSS.	
SOURCE	Drosophila melanogaster.	
ORGANISM	Drosophila melanogaster Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Peiryogeta; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila. 1 (bases 1 to 887) Genoscope. Direct Submission	
REFERENCE	Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr Web : www.genoscope.cns.fr)	
AUTHORS	Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Oosawa and Aaron Mammosses at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain Y2; cn bw sp, the same strain used for the BDGP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.	
TITLE	Location/Qualifiers	
JOURNAL	1..887 /organism="Drosophila melanogaster" /db_xref="taxon.7227" /clone="BACR26H11" /clone_id="RPCI-98" /note="end : TE13"	
COMMENT		
BASE COUNT	242 a 109 c 62 g 257 t 217 others	
ORIGIN		
Query Match	4.1%; Score 44.2; DB 17; Length 887;	
Best Local Similarity	29.2%; Pred. No. 1.8;	
Matches 202; Conservative 123; Mismatches 362; Indels 4; Gaps 1;		
Oy	AAAAAGGCAAGAAGTTAAATGATTCATTGATTAAGAGATGTTTGGCACTCAGGT 121	
Dd	181 AAAATTTTATGCATTAATAAATTAAGCTCTCCNCTAGGAAGTRCTGAATTCAGCTATCG 240	
Oy	122 CCGCATTAATAATGCAATATACCGAATTTGATTTGATTAAGTGATCTTCAAATGC 181	
Dd	241 GTTAAATAATSWCTNNATTTTTSTTTCTTTACTTAAKTTKTTTWTWSKMAT 300	
Oy	182 TTCAGATGTTTTCATCAGTAGACATTTTCCTTCATATCTGTGTTCTTTGGGNGTG 241	
Dd	301 CTCTTTTKGCCNCTTTKKKGKKKSgtgsscgctgcgttggcgcgckkmgat 360	
Oy	242 TCAAGATTTGGAGTATCATTTGTCGTGCTGTACTGCGATTTGCATAGTTTGT 301	
Dd	361 KCTTKKKKGSGITTTNGTGTGGKKKKKKKTGTNTTKGKGSBKTTCSNCGNCT 420	
Oy	302 CCTCAAGGCGCTGCTCAATGTTCCGTTTANGTTTCCCAACAATA----CTTGAGCTCA 357	
Dd	421 CSSSSSSSSSBTTTTTTTTTBSRAATTTTTTWTTMACTTATCTBTTTSSBBRYT 480	
Oy	358 AACCTTTATTTGCTCCTCACCAATGTACTCAACAGCATCTCCCTTTAGAAATGGCT 417	
Dd	481 TWTWATTAARVAAAATATCWACAYCTTWTYTTAMTAAMTTTTVAATCTTTVCYAAAY 540	
Oy	418 CAAAAGTTGGTATGAACAATTTGTTATCCAAGATGTTGCTCTTTCTAGTTCTTTCA 477	
Dd	541 TCAAATATATMACATTTTCMCATTTTWTAWAACVAAAATWTTTVCCTAAATATTTTMM 600	

Qy 478 GTTTCGATGTAATCAACAAATTAATCAATCATCATGCGATGATATCAAG 537
 Db 601 TTYCHWCAWTYTTTTTTTTTTTTTTTCTYCAAAATYGCYATTTTWCAMAMACMMAA 660
 Qy 538 GCCTGATCTTCCACATCTCGGACCAAGTCACAGATATCACTCAAAAATATAGA 597
 Db 661 ACAAYAMVCCACATYTTAAATTTTTTTTTTAAATTAATTAATTAATTAATTAATTA 720
 Qy 598 ACCATTTGATTCACAGATGTTGCTCTCAAGAGATATATTAATCTGTAAAGCAA 657
 Db 721 AAAAATTTCTTTTMMATATYCTHTHMMMAATTTTMMTTTTTAAAMAAAHWCARAC 780
 Qy 658 GGTAACCAATTTCAAACTATATGCGATATAGTAAAGTATGTGAATTAAGGAT 717
 Db 781 CTTAATTCWCTTAAAMAACTCTCTWMAHAAAAAAATATTTTTHAAAMATTTTT 840
 Qy 718 ATCATCAGATCAACATATCAATGTTCAAC 748
 Db * 841 AACCAHWCWCTCYMMHMTAAAAAACMAM 871

RESULT 5
 BG553567 518 bp mRNA linear EST 09-APR-2001
 LOCUS df03604.x1 wellcome CRC PRN3 St19 26 Xenopus laevis cDNA clone
 DEFINITION IMAGE:3550063 3', mRNA sequence.

ACCESSION BG553567
 VERSION BG553567
 KEYWORDS EST: African clawed frog.
 SOURCE Xenopus laevis
 ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipodae; Pipidae; Xenopodinae; Xenopus.

REFERENCE 1 (bases 1 to 518)
 AUTHORS Clifton,S., Johnson,S.L., Blumberg,B., Song,J., Hillier,L., Page,D., Martin,J., Wylie,T., Underwood,K., Theising,B., Bowers,Y., Person,B., Gibbons,M., Harvey,N., Ritter,E., Jackson,Y., McCann,R., Waterston,R. and Wilson,R.
 TITLE Washu Xenopus EST project, 1999
 JOURNAL Unpublished (1999)
 COMMENT Contact: Sandy Clifton, Ph.D.
 Washu Xenopus EST project, 1999
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu
 Library constructed by N. Garrett, K. Ryan and A.M. Zorn (Wellcome/CRC Institute). DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: Xenopus clones from this library are available through the I.M.A.G.E. Consortium/BLNt at: info@image.llnl.gov
 Seq primer: -400P from Gibco
 High quality sequence stop: 458.

FEATURES
 source
 1..518
 Location/Qualifiers
 1..518
 /organism="Xenopus laevis"
 /db_xref="taxon:8355"
 /clone="IMAGE:3550063"
 /clone_id="wellcome CRC PRN3 St19 26"
 /rname_type="pooled embryos, stage 19-26"
 /lab_host="DH10B (phage-resistant)"
 /note="Vector: pBSRN3, Site 1: NotI; Site 2: EcoRI; CDNA5 were oligo-dT primed and directionally cloned. Staging according to Newkirk and Faber. Library was constructed by N. Garrett, E. Bellefroid, and A.M. Zorn, (Wellcome/CRC Institute)."

BASE COUNT 197 a 87 c 97 g 137 t
 ORIGIN
 Query Match 4.1%; Score 44; DB 12; Length 518;
 Best Local Similarity 49.2%; Pred. No. 1.7;

Matches 116; Conservative 0; Mismatches 120; Indels 0; Gaps 0;
 Qy 628 CAAGAGAAATATTAATCTGTAAAGCAAGTAAACCAATTTCAACTATGTCGA 687
 Db 107 CAAAGAAATATTAATCTGTAAAGCAAGTAAACCAATTTCAACTATGTCGA 166
 Qy 688 TATAGGAAAGTATGTCGAATTAAGATATCATAGAGATCAACATTAATGTTCA 747
 Db 167 TATAGCACTGTAAAGTGTATTAAGAAAAACAAAAAGCTTAATGTGATCA 226
 Qy 748 CTCATGTTACCTCTGTATTAAGCATAGTTATGATGTTTGGCTTGAATTCAT 807
 Db 227 GTCAAAAGAGTCTTTCCCACTGTTGTTGTTTAAATTTCAATTTCAAT 286
 Qy 808 TTGTGATGAAAGCAAAAGTGTGTTGAATTAAGATTAATTAATGACTAA 863
 Db 287 TTTTCTTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 342

RESULT 6
 CNS00601 990 bp DNA linear GSS 03-JUN-1999
 LOCUS BACR14023 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
 DEFINITION AL065624
 ACCESSION AL065624
 VERSION AL065624
 KEYWORDS GSS: Drosophila melanogaster.
 SOURCE Drosophila melanogaster.
 ORGANISM Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE 1 (bases 1 to 990)
 AUTHORS Direct Submission
 TITLE Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : sequef@genoscope.cns.fr - Web : www.genoscope.cns.fr)
 JOURNAL Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mamoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw bp, the same strain used for the BDGP's pl and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES
 source
 1..990
 Location/Qualifiers
 1..990
 /organism="Drosophila melanogaster"
 /db_xref="taxon:7227"
 /clone="BACR14023"
 /clone_id="RPCI-98"
 /note="end : TET3"

BASE COUNT 158 a 51 c 40 g 592 t 149 others
 ORIGIN
 Query Match 4.1%; Score 44; DB 17; Length 990;
 Best Local Similarity 36.9%; Pred. No. 2.1; 395; Indels 2; Gaps 1;
 Matches 262; Conservative 51; Mismatches 395; Indels 2; Gaps 1;

Qy 189 TGTTCCTCAAGATGATGATTTCTTCATATCTGTTCTTGGGCTGCTCAAGAG 248
 Db 190 TTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTT 249
 Qy 249 TTGATGAATCATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 308

Db 250 TTTT... 309
 QY 309 TGCCTGTCAATGCTCCGTTATAGTTCCCAATACCTGCTCAACCTTTAT 368
 Db 310 TTTT... 369
 QY 369 GTCTCAACCAATGTATCAACAGCATCTCCCTTTAGAAATGGCTCAAAAGTTGG 428
 Db 370 TTTT... 429
 QY 429 TTTAAGCAAAATGTTATACAGATGTTGCTCTTCAAGTTCTTCAAGTTGGATT 488
 Db 430 TTTT... 489
 QY 489 GTATCAACATTAATCTCAACATCATCATGAGTATGATATCAAGGCTGTGATC 548
 Db 490 TTTT... 549
 QY 549 TCCACAATCTGTGACCAAGTCAACAGATATCATCTCAAAATATGAAACATTTTGA 608
 Db 550 TATTT... 609
 QY 609 TTTCAACAGTTGCTCTCTCAAGAGAAATATTAATCTGTGAAGCAAGTAAACCAAT 668
 Db 610 TCGAMTMTAARRTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTT 669
 QY 669 TTTCAACTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 728
 Db 670 TTTT... 729
 QY 729 TCAACATATCAATGATCACT--CATGTATCCTGTATGATGATGATGATGATGAT 786
 Db 730 TTTATATATATTTT... 789
 QY 787 GTTTGGCTTTGAATTCATTTGTGTGATGATGATGATGATGATGATGATGATG 846
 Db 790 TTTATATATATTTT... 849
 QY 847 AAGAATTAATGATTAATGATTTGATGATTTGATGATTTGATGATTTGATGAT 896
 Db 850 TTTT... 899

RESULT 7
 CNS00EVL/c 1101 bp DNA linear GSS 04-JUN-1999
 LOCUS 1101 bp DNA linear GSS 04-JUN-1999
 DEFINITION Drosophila melanogaster genome survey sequence 77 end of BAC:
 BACR29823 of RPECI-98 library from Drosophila melanogaster (fruit
 fly), genomic survey sequence.
 ACCESSION AL069706
 VERSION AL069706.1 GI:4949849
 KEYWORDS GSS.
 SOURCE Drosophila melanogaster.
 ORGANISM Drosophila melanogaster.
 Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 Ephydroidea; Drosophilidae; Drosophila.
 1 (bases 1 to 1101)
 Genoscope.
 Direct Submission
 Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
 BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
 - Web : www.genoscope.cns.fr)
 Determination of this BAC-end sequence was carried out as part of a
 collaboration with the Berkeley Drosophila Genome Project (BDGP).
 The BDGP is constructing a physical map of the Drosophila
 melanogaster genome using these BACs. For further information
 please see <http://www.fruitfly.org> The BDGP Drosophila
 melanogaster BAC library was prepared by Kazuo Osoegawa and
 Aaron Hammar in Pieter de Jong's laboratory in the Department of
 Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
 NY. The library is named RPECI-98 and was constructed by partial

Scori digestion of Drosophila DNA provided by the BDGP from the
 isogenic strain Y2; cn bw sp, the same strain used for the BDGP's
 pl and EST libraries. A more detailed description of the library
 and how to order individual BAC clones, the entire library, or
 filters for hybridization from the BACPAC Resource Center can be
 found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES
 source
 1. 1101
 /organism="Drosophila melanogaster"
 /db_xref="taxon:7227"
 /clone="BACR29823"
 /clone_1b="RPECI-98"
 /note="end : 17"
 BASE COUNT 419 a 91 c 60 g 299 t 232 others
 ORIGIN

Query Match 4.1%; Score 44; DB 17; Length 1101;
 Best Local Similarity 29.4%; Fred. No. 2.2; Mismatches 0; Gaps 0;
 Matches 178; Conservative 103; Mismatches 325; Indels 0;

QY 101 ATGTTTGCACAGCTCAAGTCTCCGATATATATATCAATATGATGATGATGATGAT 160
 Db 1047 AATTTATATATATTTATTTTAAATATATATATATATATATATATATATATAT 988
 QY 161 GTGATGCTACTCTCAAGTCTTCAGATGTTTCTCACTGATGATGATGATGATGAT 220
 Db 987 MATMAMAMTATAMATMAMATMAMATMAMATMAMATMAMATMAMATMAMATTTT 928
 QY 221 ATCTGTTTCTGGGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 280
 Db 927 TTTT... 868
 QY 281 GTGATGCTACTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 340
 Db 867 ATTATATATATATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTAT 808
 QY 341 CACAATATGTTGCTCAAACTTTTATTTGCTCAACCAATGTTGTTGTTGTTGTTG 400
 Db 807 TATTTT... 748
 QY 401 CCGTTTATGATGCTCAAAAGTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 460
 Db 747 MATTTT... 688
 QY 461 CTTCTGTTCTTCAAGTCTTCGATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 520
 Db 687 TTTT... 628
 QY 521 CTGAGTGTATCAAGGCTCTGATCTTCACATCTGTTGTTGTTGTTGTTGTTGTT 580
 Db 627 TATMAAAATATTTT... 568
 QY 581 TCACTCAAAATATAGAACATTTTGATTTCAACAGTTGTTGTTGTTGTTGTTGTT 640
 Db 567 TATTTT... 508
 QY 641 TTAATCTGTAAGCAAGGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 700
 Db 507 TTTT... 448
 QY 701 TGTGTG 706
 Db 447 KKKKKG 442

RESULT 8
 CNS01219/c 1101 bp DNA linear GSS 26-JUL-1999
 LOCUS 1101 bp DNA linear GSS 26-JUL-1999
 DEFINITION Drosophila melanogaster genome survey sequence SP6 end of BAC:
 BACR08024 of DrosBAC library from Drosophila melanogaster (fruit
 fly), genomic survey sequence.
 ACCESSION AL101595
 VERSION AL101595.1 GI:5613206

KEYWORDS GSS.
SOURCE Drosophila melanogaster.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1101)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a
collaboration with the European Drosophila Genome Project (EDGP) -
http://www.edgp.ebi.ac.uk - This Drosophila melanogaster BAC
library (Dros BAC) was made by Alain Billand at CEPH (Centre
d'Etude du Polymorphisme Humain) with funding provided by a MRC
project grant. The DNA was prepared from embryos by Alain Bucheton
and Genevieve Payan. It has been constructed in the vector
pBelobAC11.
FEATURES
source
1..1101
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_1lb="BACN08024"
/clone_1lb="DrosBAC"
/plasmid="pBelobAC11"
/note="end : SP6"
BASE COUNT 436 a 142 c 169 g 227 t 127 others
ORIGIN
Query Match 4.1%; Score 44; DB 17; Length 1101;
Best Local Similarity 36.4%; Pred. No. 2.2;
Matches 134; Conservative 42; Mismatches 192; Indels 0; Gaps 0;
QY 127 TATATATCGAATATACGAATTGATGTTACAGTATGCTAGCTTCAAGCTTCCA 186
DB 564 TTTTATATTAATATACGGGTTTAACTTAATTAATTAATTAATTAATTAATA 505
QY 187 GATGTTTTCACAGTATGATGATGCTTCCATATCTGTTTGGGGGTCAAG 246
DB 504 TTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 445
QY 247 AGTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 306
DB 444 ATTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 385
QY 307 AGTCCCTGTGCAATGCTCCGTTTATGTTTCCACAAATCTGTGCTCAACTTTTA 366
DB 384 TTTTATCCCKKTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTT 325
QY 367 TTGCTCACAAGAGTCAACAAGATCTCCCTTTAGATGATGCTCAAAAGGT 426
DB 324 TTTTAAATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 265
QY 427 GGTATGAACAATTTGTTATTCAGATGTTGCTCTTCTAGTCTTTCACAGTTTTCGA 486
DB 264 TTTAMKWTATMCTTTTCCCTTTTATTTTATTTTATTTTATTTTATTTTATTTT 205
QY 487 TTGATCA 494
DB 204 TTTTWTTA 197
RESULT 9
AM774971
LOCUS 520 bp mRNA linear EST 07-SEP-2000
DEFINITION ESTJ34122 KV3 Medicago truncatula cDNA clone pKV3-25E23, mRNA
sequence.
ACCESSION AM774971
VERSION AM774971.1 GI:7718888
KEYWORDS EST.
SOURCE batrel medic.

ORGANISM Medicago truncatula
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
Medicago.
REFERENCE 1 (bases 1 to 520)
AUTHORS VandenBosch, K., Hurt, J., Moore, J., Beremand, P., Peng, H., Ellis, L.,
Town, C.D., Bowman, C.L., Craven, M.B., Hansen, T.S., Holt, I.E. and
Fraser, C.M.
TITLE ESTs from roots of Medicago truncatula after Rhizobium inoculation
JOURNAL Unpublished (1999)
COMMENT Contact: VandenBosch K
Department of Plant Biology
University of Minnesota
220 Biosci Center, 1445 Gortner Ave, St. Paul, MN 55108, USA
Tel: 612 624 2755
Fax: 612 625 1738
Email: kvandenb@chs.umn.edu
Texas A&M EST name: MTEBK367K
TIGR sequence name: MTEBK367K
More information is available at:
http://chryslie.tamu.edu/medicago
Seq primer: Skmod (CTA gAA CTA gTg gAT CC).
FEATURES
source
1..520
/organism="Medicago truncatula"
/cultivar="genotype A17"
/db_xref="taxon:3880"
/clone_1lb="pKV3-25E23"
/clone_1lb="KV3"
/tissue_type="Seedling roots"
/dev_stage="3 days post-inoculation with Sinorhizobium
meliloti"
/lab_host="E. coli strain XL0LR"
/note="Vector: pBluescript SK -; Site 1: EcoRI; Site 2:
XhoI; cDNA was prepared from polyA+ enriched RNA. The cDNA
was directionally ligated into the UniZap XR vector from
Stratagene and packaged using Gigapack III Gold packaging
extracts. Plasmids containing cDNA inserts were excised
from the recombinant lambda-Zap phage using Ex-assist
helper phage and propagated in XL0LR cells."
BASE COUNT 188 a 92 c 110 g 130 t
ORIGIN
Query Match 4.1%; Score 43.8; DB 10; Length 520;
Best Local Similarity 49.0%; Pred. No. 1.9;
Matches 117; Conservative 0; Mismatches 122; Indels 0; Gaps 0;
QY 620 TTGCTCTCAAGAGATATTAATCTCTGTAAGCAAGTAAACAATTTCAACTACT 679
DB 247 TTGCTGCTCAAAAAGCAGCTGCATTTGCTTAACAGCAAAACAAGCTGCTTCAAAAACA 306
QY 680 ATGCGGATATAGTAAGATATGTAATTAAGATATATCATGAGATCAACAATACA 739
DB 307 ATGTCACAGAAAAGAGAGATATGTAATAAACAATGATGCGAAGAAATTCACAAGATG 366
QY 740 ATGTTCAACTGATGTAACCTGATTAAGCAATACGTTTACTGATGTTTGGCTTTGA 799
DB 367 AAGTGTGTCGAAGAAGATCGAGATGTAAGAAAGTTTGAAGTCATGCTTTGATTCAA 426
QY 800 ATTGCAATTTGTGTGATGAAGCAAAAGGTGCTGTGTAAGATTATCAAGATTATTG 858
DB 427 ATATGAGAAATGATATGATGATGAAGTAACAAGTTAGAAAGATGCGACACAATAG 485
RESULT 10
BM276144
LOCUS 527 bp mRNA linear EST 20-DEC-2001
DEFINITION PEST00a70h09.v1 Plasmodium falciptarum 3D7 gametocyte cDNA library
Plasmodium falciptarum 3D7 cDNA 5', mRNA sequence.
ACCESSION BM276144
VERSION BM276144.1 GI:17969494
KEYWORDS EST.

SOURCE Plasmodium falciparum 3D7.
ORGANISM Plasmodium falciparum 3D7
REFERENCE Eukaryote; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
AUTHORS 1 (bases 1 to 527)
Tang, K., Cole, R., Chakrabarti, D., Haywood, R., Clifton, S., Page, D.,
Matta, M., Hillier, L., Martin, J., Wylie, T., Dante, M., Theising, B.,
Bowers, Y., Gibbons, M., Ritzer, E., Bennett, J., Jentes, E., Ronko, I.,
Teagareishvili, R., Belagorod, L., Franklin, C., Carr, L., Grow, A.,
Maguire, L., Richey, J., Madkins, J., Kennedy, S., Levinso, D.,
Waterston, R., Wilson, R. and Sibley, D.

TITLE MASHU Plasmodium EST Project
JOURNAL Unpublished (2001)
COMMENT Contact: L. David Sibley
MASHU Plasmodium EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu

Library was constructed by R. Haywood. DNA sequencing by:
Washington University Genome Sequencing Center For information on
obtaining a clone please contact: L. David Sibley
(sibley@borcim.wustl.edu), Washington University
Seq primer: -40UP from Gibco
High quality sequence stop: 423.
Location/Qualifiers

FEATURES

source

1. 527
/organism="Plasmodium falciparum 3D7"
/db_xref="taxon:36329"
/clone_id="Plasmodium falciparum 3D7 gametocyte cDNA
library"
/dev_stage="gametocyte (stage III-V)"
/lab_host="DH10B (Genehogg, Invitrogen, Inc.)"
/note="Vector: pBluescript SK plus; Site 1: EcoRI; Site 2:
XhoI. The library was constructed by R. Haywood. cDNAs were
synthesized from gametocyte poly(A) + RNA by oligo d(T)
priming, size-selected and directionally cloned into the
EcoRI (5' end) to XhoI (3' end) sites of the Uni-ZAP XR
lambda vector (Stratagene). The primary library was mass
excised as phagemid using the Exassist helper phage
(Stratagene). Clones were mass excised using the Exassist
helper phage (Stratagene), the phagemids were precipitated
with PEG 8000 and extracted with phenol/chloroform.
Phagemid DNA was electroporated into DH10B cells. Clone
Availability: David Sibley, Washington University."

BASE COUNT 243 a 57 c 88 g 139 t

ORIGIN

Query Match 4.1%; Score 43.6; DB 13; Length 527;
Best Local Similarity 48.8%; Pred. No. 2.2;
Matches 118; Conservative 0; Mismatches 124; Indels 0; Gaps 0;

QY 807 TTGTGATGAAAGCAAAAGTTGCTGTTGAAGTATCAAGATTATGACTAATAC 866
DB 55 TTGTGATGAAAGCAAAAGTTGCTGTTGAAGTATCAAGATTATGACTAATAC 114

QY 867 TTGTGATGAAAGCAAAAGTTGCTGTTGAAGTATCAAGATTATGACTAATAC 926
DB 115 ATCATATATGACCAAAATGATGCGGCTGCAATTTGCGACCAACATGATGTTATTGA 174

QY 927 TCATTGGCTAAATCATCAACTTTTATGCTCAATTGAGCCAAATTCGATGCAAGA 986
DB 175 AGATAGCTGCTCATCTCATCATGATCATGAATATATAGAAACAAATTAAGA 234

QY 987 AAGTGAAGTTAGACTTTTGAAGATGTTGACTTTGCTAACAGAACTTAAGATTGTC 1046
DB 235 AAAACCGAAGCAGTAAAGAGTGAATTTGTATATATGATTTAAAAAAAATTGA 294

QY 1047 TG 1048
DB 295 TG 296

RESULT 11 530 bp mRNA linear EST 20-DEC-2001
LOCUS BM273621
DEFINITION PFE870a57c09.y1 Plasmodium falciparum 3D7 gametocyte cDNA library
Plasmodium falciparum 3D7 cDNA 5', mRNA sequence.

ACCESSION BM273621 GI:17966918
VERSION BM273621.1
KEYWORDS EST.
SOURCE Plasmodium falciparum 3D7.
ORGANISM Plasmodium falciparum 3D7.
Eukaryote; Alveolata; Apicomplexa; Haemosporida; Plasmodium.

REFERENCE 1 (bases 1 to 530)
Tang, K., Cole, R., Chakrabarti, D., Haywood, R., Clifton, S., Page, D.,
Matta, M., Hillier, L., Martin, J., Wylie, T., Dante, M., Theising, B.,
Bowers, Y., Gibbons, M., Ritzer, E., Bennett, J., Jentes, E., Ronko, I.,
Teagareishvili, R., Belagorod, L., Franklin, C., Carr, L., Grow, A.,
Maguire, L., Richey, J., Madkins, J., Kennedy, S., Levinso, D.,
Waterston, R., Wilson, R. and Sibley, D.

TITLE MASHU Plasmodium EST Project
JOURNAL Unpublished (2001)
COMMENT Contact: L. David Sibley
MASHU Plasmodium EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu

Library was constructed by R. Haywood. DNA sequencing by:
Washington University Genome Sequencing Center For information on
obtaining a clone please contact: L. David Sibley
(sibley@borcim.wustl.edu), Washington University
Seq primer: -40UP from Gibco
High quality sequence stop: 422.
Location/Qualifiers

FEATURES

source

1. 530
/organism="Plasmodium falciparum 3D7"
/db_xref="taxon:36329"
/clone_id="Plasmodium falciparum 3D7 gametocyte cDNA
library"
/dev_stage="gametocyte (stage III-V)"
/lab_host="DH10B (Genehogg, Invitrogen, Inc.)"
/note="Vector: pBluescript SK plus; Site 1: EcoRI; Site 2:
XhoI. The library was constructed by R. Haywood. cDNAs were
synthesized from gametocyte poly(A) + RNA by oligo d(T)
priming, size-selected and directionally cloned into the
EcoRI (5' end) to XhoI (3' end) sites of the Uni-ZAP XR
lambda vector (Stratagene). The primary library was mass
excised as phagemid using the Exassist helper phage
(Stratagene). Clones were mass excised using the Exassist
helper phage (Stratagene), the phagemids were precipitated
with PEG 8000 and extracted with phenol/chloroform.
Phagemid DNA was electroporated into DH10B cells. Clone
Availability: David Sibley, Washington University."

BASE COUNT 246 a 57 c 88 g 139 t

ORIGIN

Query Match 4.1%; Score 43.6; DB 13; Length 530;
Best Local Similarity 48.8%; Pred. No. 2.2;
Matches 118; Conservative 0; Mismatches 124; Indels 0; Gaps 0;

QY 807 TTGTGATGAAAGCAAAAGTTGCTGTTGAAGTATCAAGATTATGACTAATAC 866
DB 55 TTGTGATGAAAGCAAAATGATGATGATATTAATCAAAATGAAATACCATTAATCC 114

QY 867 TTGTGATGAAAGCAAAAGTTGCTGTTGAAGTATCAAGATTATGACTAATAC 926
DB 115 ATCATATATGACCAAAATGATGCGGCTGCAATTTGCGACCAACATGATGTTATTGA 174

QY 927 TCATTGGCTAAATCATCAACTTTTATGCTCAATTGAGCCAAATTCGATGCAAGA 986
DB 175 AGATAGCTGCTCATCTCATCATGATCATGAATATATAGAAACAAATTAAGA 234

QY 987 AAGTGAAGTTAGACTTTTGAAGATGTTGACTTTGCTAACAGAACTTAAGATTGTC 1046

Db	235	AAACCAAGCAGTAAAGACTGATGTGATATCATTTAAAAAAAATTGA	294
Qy	1047 TG 1048		
Db	295 TG 296		

RESULT 12					
BM276328		513 bp	mRNA	linear	EST 20-DEC-2001
LOCUS	BM276328				
DEFINITION	Pf3STGa8d1d1.y1 Plasmodium falciparum 3D7 gametocyte cDNA library				
	Plasmodium falciparum 3D7 cDNA 5', mRNA sequence.				

SOURCE *Plasmodium falciparum* 3D7.
ORGANISM *Plasmodium falciparum* 3D7.
REFERENCE Eukaryotic Alveolata; Apicomplexa; Haemosporida; *Plasmodium*
1 (bases 1 to 533)
AUTHORS Tang, R., Cole, R., Chakrabarti, D., Haywood, R., Clifton, S., Pe
Marr, M., Hillier, L., Martin, J., Willis, T., Dane, M., Theis
Bowyer, Y., Gibbons, M., Ritter, E., Bennett, J., Jentes, E., Ro
Tsegareishvili, R., Belagorod, L., Franklin, C., Carr, L., Gro
Maguire, L., Richey, J., Wadkins, J., Kennedy, S., Levinso, D.,
Waterson, R., Wilson, R. and Sibley, D.
TITLE Mashu *Plasmodium* EST Project
JOURNAL Unpublished (2001)
COMMENT Contact: L. David Sibley

BASE COUNT	249 a	57 c	88 g	139 t
ORIGIN				

	Query Match	Similarity	Score	DB	Length
Best Local	18%	48.8%	43.6	13	533
Matches	118	conservative	0	Mismatches	124
				Indels	0
				Gaps	0

	Query	Match	Similarity	Score	DB	Length
807	TTTTGCTGATGAAAGCAAAAGCTGCTGTGAATATCAAGAAATTTATGATAATAC	866	86	86	86	86
55	TTTGTTAAGTGAATATAAATATGATATGATATTAATTAACAAATGAATATCAATCAATCC	114	114	114	114	114
867	TTTAGTTTGGACTGTGTGGGCTCGATTAATCTCCAGGCAACAAGAAATGGATTGC	928	928	928	928	928

Db	115	ATCATATATGACCAAAATGATGGGGCTGCATTTTCGACCCCAAGATGATTATTTGA	174
Qy	927	TCATTTGGCTAAATCATCAAACTTTTATGCTCAATTGAGCCAACTTGATGCCAAGGA	986
Db	175	AGATACGTGCTCATCACTCTCATCAGATCATGAATATTAAGAAAGAAAACAAATTAAGA	234
Qy	987	AAGTGAAGTTAGAGTTTGGATGTGTGTCCTTTCCTAAACAGGAAGTTAAGAAATTGTGC	1046
Db	235	AAAACCGAAGCAGTAAAGAGTGAAGTATTTGTATTAATGATTTAAAAAATAATTGA	294
Qy	1047	TG 1048	
Db	295	TG 296	

RESULT 13	
BM275241	
LOCUS	570 bp mRNA linear EST 20-DEC-2001
DEFINITION	Plasmodium falciparum 3D7 gametocyte cDNA library
ACCESSION	BM275241
VERSION	BM275241.1 GI:17968565
KEYWORDS	EST.
SOURCE	Plasmodium falciparum 3D7.
ORGANISM	Plasmodium falciparum 3D7.
REFERENCE	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
AUTHORS	1 (bases 1 to 570) Tang, K., Cole, R., Chakrabarti, D., Hayward, R., Clifton, S., Page, D., Datta, M., Theisen, B.

Unpublished (2001)
 Contact: L. David Sibley
 Washu Plasmodium EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu
 Library was constructed by R. Haywood. DNA sequencing by:
 Washington University Genome Sequencing Center For information on
 obtaining a clone please contact: L. David Sibley
 (sibley@orcim.wustl.edu), Washington University
 Seg primer: -40UP from Gibco
 High quality sequence stop: 424.

```

FEATURES
    source
        Location/Qualifiers
            1..570
                /organism="Plasmodium falciparum 3D7"
                /db_xref="taxon:36329"
                /clone_lib="Plasmodium falciparum 3D7 gametocyte cDNA
                    library"
                /dev_stage="gametocyte (stage III-V)"
                /lab_host="DH10B (GeneHog, Invitrogen, Inc.)"
                /note="The library was constructed by R Haywood. cDNAs were
                    synthesized from gametocyte poly(A) + RNA by oligo d(T)
                    priming, size-selected and directionally cloned into the
                    EcoRI (5' end) to XhoI (3' end) sites of the Uni-ZAP XR
                    lambda vector (Stratagene). The primary library was mass
                    excised as phagemid using the Exassist helper phage
                    (Stratagene). Clones were mass excised using the Exassist
                    helper phage (Stratagene), the phagemids were precipitated
                    with PEG 8000 and extracted with phenol/chloroform.
                    Phagemid DNA was electroporated into DH10B cells. Clone
                    Availability: David Sibley, Washington University."
    267 a
        61 c
            93 g
                149 t

```

Matches	Conservative	Mismatch	Indels	Gaps
363	151	218	6	422

Page 9

